DOCUMENT RESUME

ED 268 149 TM 860 140

AUTHOR	Gulliksen, Harold
TITLE	Creating Better Classroom Tests.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RM-85-5
PUB DATE	Dec 85
NOTE	72p.
PUB TYPE	Viewpoints (120) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Adults; Classroom Techniques; *Educational Testing; Educational Trends; Multiple Choice Tests; *Teacher Made Tests; Teachers; *Test Construction; *Test Format; *Testing Problems; *Test Items; Test Manuals; Test Selection

ABSTRACT

This article presents the perspective that the quality of teacher-made, small classroom tests has not improved, and may have declined in recent years. This decline may be due to the fact that teachers have come to believe that the kinds of objective items used in national standardized tests are the only item types appropriate for classroom use. The points stressed for improving classroom tests include: (1) teachers should use the item form that is most natural and convenient for a particular academic subject, instead of predominantly relying on the multiple-choice item test; (2) teachers should learn the usefulness of performance and identification items when writing items, especially in courses that include laboratory work; and (3) teachers should use the improved design of free-answer items (that is, medical diagnosis items, in-basket items, etc). The extensive appendices include sample test items and illustrative item types. Also attached are two previously published articles: "Measuring the Effects of Newspapers in the Classroom," by Paul B. Diederich and Marvin Maskovsky; and "Evaluating Innovations by Innovations in Evaluation," by Paul B. Diederich. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED268149

RESEARCH MEMORANDUM

CREATING BETTER CLASSROOM TESTS

Harold Gulliksen

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. Weidenmiller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Educational Testing Service
Princeton, New Jersey
December 1985

TM 860 140

Copyright © 1985. Educational Testing Service. All rights reserved.

I think it is important for persons interested in education, especially learning, testing, and evaluation, to bear in mind that large-scale standardized testing programs require that a test or tests be given to thousands of students only a few times a year using standardized machine-scored answer sheets. Such instruments are designed to test objectives common to a large number of different schools. The foregoing situation must be differentiated from the requirements of and the methods appropriate for tests prepared by the individual teacher, given on a weekly or monthly basis and covering work on the last chapter or on a segment of the course studied by a particular class. Such tests should be scored and returned to the students in a day or, in some cases, for short tests given at the beginning of the class period, scored by students swapping papers and returned in five minutes for class discussion. The requirements of, and the methods appropriate for, these two types of tests and testing situations are very different. I am beginning to believe that the failure to make this distinction is responsbile for there having been no improvement, and perhaps even a decline, in the quality of teacher-made, small classroom tests over the last forty years. This decline may well be due to the fact that teachers have come to believe that the kinds of objective items used in national standardized tests are the only item types appropriate for classroom use. Where such items are not satisfactory, then the only alternative is the essay test. To give teachers the idea that using computers and machine scoring for classroom tests is desirable, or that machine-scorable item types are the best or the only ones to use, has and will inhibit progress.

We should note that there are over 40 million public school students in the United States, and as of a year or two ago about 600,000 or 700,000 computers. The important thing to stress now is what teachers can do without computers. Those who have and use computers could be given special work on how to use them in testing, provided that instructors could be found who are sufficiently well-informed on the use of computers.

Certain developments in large-scale standardized testing, if used in classroom testing of a recent assignment, make such testing very difficult and laborious.

One such difficulty is the use of computer answer sheets and computer scoring. Giving and scoring a test in the opening ten minutes of a class period does not allow the use of a computer, unless each student in the class has his own personal computer or hand-held scoring device. This situation will not occur for a very long time yet, and teachers should not be given the impression that they should wait until each student has a computer before starting to use classroom tests. Similarly, the use of computers or computer scored answer sheets is a handicap. Sometimes the answer sheets are not available for a given class, and they restrict the item forms that can be used. Having provision for the answer on the examination means that getting the exam mimeographed or xeroxed is all that is necessary. Separate answer sheets, and computerized answer sheets that are a necessary convenience in large-scale standardized testing, are an inconvenience in small-scale classroom testing.

I want to emphasize that I am not referring to the use of computer answer sheets and computer scoring of answer sheets for large-scale

testing programs. There computers are essential. However, for an individual class of 30 to 50 or so students, the use of computers will

1. Limit the item types to be used;
2. Cause delay and inconvenience by waiting for answer sheets to arrive, or for machine time to be available; and
3. Cause delay for return of papers by machine scoring.

For classroom use a variety of item types adjusted according to subject matter are appropriate. The computer discourages or prevents this tailoring. Note that I am in favor of and not objecting to the use of objective items in classroom tests - but only pointing out the difficulties when separate machine-scored answer sheets are used in the classroom.

One device I have found very effective in stimulating class discussion is to open the class with a ten-minute test. It can be immediately scored by the students. After the scoring, I have never had to say, "Are there any questions or comments?" They come immediately and result in an excellent class discussion period. This cannot be done using computers unless each student in the class has his own.

Another modern trend in large-scale standardized testing is toward the practically exclusive use of five alternative multiple-choice items. Textbooks of the 60's and 70's say "There are three types of items, essay, true-false, and multiple-choice," and then devote a chapter, or so to each one. Texts of the 30's and 40's list and discuss many other types of objective items that are more appropriate than multiple-choice items for many types of questions. These texts also describe several types of free-answer items other than the essay, such as the brief

answer; or the comment-on-the-following-statement; or the interlinear; or the fill-in-the-blanks-in-the-following-paragraph. All are very useful types of free-answer item, especially for the individual teacher with small classes.

Test developers encourage the belief that the principal item type is the five-alternative, multiple-choice item. Ben Bloom, for instance, devotes space to illustrating how an item that is naturally written as a nine-choice item can be transformed into five-choice item just by not marking four of the alternatives, and not presenting any statements that have any of these four choices for the answer, a procedure that clearly damages the question.

Teacher-made small-class tests should be used frequently, for instance when the class has finished a chapter of a book, or a segment of a course, to make clear to the students what they should have learned in this part of the course. It could also be a good idea to give such a test before the class studies the material, partly to see how much of it the students already know, and partly to give them a clear-cut idea of what they are supposed to be learning. In the case of individualized instruction, where each student proceeds at his own pace, the same type of pretest and post-test procedure would be valuable.

Such tests must be constructed by the teacher who is dealing with a particular class assignment. A set of test items that can be purchased will not meet the needs of the better teachers. Searching through a stack of items prepared by someone else is a very inefficient method of constructing a test for a particular segment of an individual course. A

book of sample items that teachers might survey would be very helpful in suggesting types of items that might be used.

When I came to Princeton in 1942 to work at the College Entrance Examination Board for John Stalnaker, there was considerable discussion about the necessity for teaching teachers how to write good exams for their classes. This training would cover both objective and free-answer exams, as appropriate, and place emphasis on clearly stated items and reliable scoring of the free-answer items. As far as I can make out by talking to persons acquainted with teacher training and performance and comparing texts on item construction written in the 60's and 70's with those written in the 30's and 40's, and by looking at texts that have booklets of test items with them (of which there were none in the 30's and 40's), the situation now is worse than it was in the 40's. Now we have texts accompanied by a booklet of rather poor multiple-choice items, whereas 40 years ago there were no such booklets.

Looking at the quality of exam items frequently given as supplements to texts, and hearing some of the persons at ETS who were teaching part time say "In my own course I do not use objective items; I use essay items because I want to know what the students understand," I have often thought that over the last 35-40 years the situation regarding teaching teachers to write good exam items for their classes has become worse, rather than better.

I mentioned this to Bill Turnbull, a few years before he retired from the ETS presidency. He agreed with my appraisal, thought that something should be done, and set up a committee under John Helmick to prepare a manual on item writing that could be used by teachers interested in writing better test items.

Some members of the committee were baffled by the strange restriction that was being imposed on them: they were not to write multiple-choice items. Progress on the manual was rather slow, and one of the committee members, Sybil Carlson, took over the job. With the help of a number of teachers in nearby schools, she prepared a handbook illustrating many objective item types, other than multiple-choice, for use in various subject-matter fields. This handbook has been published by ETS as "Creative classroom testing: Ten designs for assessment and instruction" and will be a helpful guide to teachers who are beginning to write objective items for their classes.

Having sketched in my general perspective, I will review briefly the major experiences of the past fifty years that have brought me to this point of view.

About 1930, President Robert Maynard Hutchins introduced an Examination System at the University of Chicago. The curriculum for freshman and sophomore years was five one-year courses: Biological Science, Physical Science, Social Science, Humanities, and English. Passing each of these courses consisted in passing a six-hour exam given in June (three hours morning, and three hours afternoon). Those who failed the June exam did not need to wait a year to take it again; it was offered in September as well.

Initially in 1930 Thurstone was appointed Chief Examiner, Marion Richardson was examiner in physical sciences, James Thomas Russell, in biological sciences, John Stalnaker in humanities and English, (Russell and Stalnaker were Examiners from 1931-1936, Harold Gulliksen, Examiner in Social Sciences, 1934-1940.) Later, Dael Lee Wolfle was Examiner in biological sciences, George Frederic Kuder, and Dorothy Adkins were also examiners. I should mention here that in 1947 when Dorothy Adkins had gone to work for the Civil Service Commission she and others prepared a very good book ("Construction and analysis of achievement tests" by Dorothy C. Adkins, Ernest S. Primoff, Harold L. McAdoo, Claude F. Bridges, and Bertram Fowler) on constructing objective and performance tests. It is in the bibliography available here for those interested.

Ralph Tyler replaced Thurstone as Chief Examiner in the late 30's and Ben Bloom was associated with the examining office later. One of the first rules Thurstone established was "The day after an exam is given it goes on sale in the University of Chicago Bookstore. The examiners and teaching faculty at Chicago together constructed the examination items, which were then reviewed, revised, and approved by the teaching staff in the course. The last item in the morning half, and also in the afternoon half was a one- or two-page essay. Short-answer items, two or three to the page ("put a word in the blank," "correct the bracketted portions of the paragraph," etc.), were also included. The rest of the written exam was objective items of various types ("master list", "true-false," "greater-less-same," "best-worst answer," "rank order," "tabular [matrix]", etc.). The bulk of the exam consisted of objective items. In many cases it was possible to construct items with a mutually exclusive and exhaustive

list of all possible alternatives (e.g., "increases," "decreases," "stay the same," "not enough information to determine"). Thus it was not necessary to construct plausible distractors for each question.

The exams at Chicago were not composed entirely of written items. Where laboratory work was important, as in biological and physical sciences the final exam involved laboratory setups. We noticed that in the laboratory class a fair amount of the instructor's time was spent moving around the laboratory and pointing out to students various things that were wrong with the apparatus setup they were using. The students corrected these errors when the instructor pointed them out. So for the exam we set up a number of laboratory experiments and had the students look at the setup and write down what was wrong or indicate that the setup was proper and ready to use. Also several of the stations would involve the actual performance of an experiment, or part of an experiment, such as dissecting a frog's leg for a nerve muscle preparation. A laboratory assistant was present at each station to give general directions as to what was to be done, and to grade the performance. Instructions to the lab assistant as to exactly what to say, and the points to watch for in grading were given beforehand. Similarly in physics and chemistry, apparatus setups would be criticized, and an experiment or two performed as part of the final examination, in addition to the written exam.

These test items proved to be excellent teaching and learning devices, helping to make it clear to the students just what they were expected to learn to do in the laboratory work. Computers could be used for such simulations, having the student make corrections by making the

appropriate keyboard entries. The point to be stressed, however, is that the computer is not necessary for such items. A regular lab experiment can be set up for each member of the class, even when that many computers are not available, and the actual equipment may give a more realistic item than the computer would.

Part of the teacher training in schools of education should be in writing items of these various types, including free-answer, objective, and also performance and identification items, emphasizing the distinctions and the skills that the teacher is trying to teach.

I still remember a very good course and exam in radio that I had in high school. A month or two before the end of the course, the instructor handed out a list of one or two hundred questions and said, "Your final exam will be ten of these." We all worked hard getting the answers to the questions, and as a result, learned what the instructor wanted us to learn in the course. Similarly, giving the students a large number of items and saying, "The final exam will be fifty or a hundred of these" would be a good teaching procedure. The teacher should write the items over a period of time, and not be handed them from an outside source. For some items, such as translating from or into a foreign language, or interpreting graphs, the same words, or graphs need not be used in the final, but a parallel item of the same type. The illustrative items should change as the emphasis and content of the course changes.

Returning to the work at the University of Chicago examining office, we checked the reliability of the total exam by correlating the morning and afternoon scores. The various parts of the exam were also checked by correlating the two essays, parallel content pages, etc. Short-answer items, with answers, were saved, and objective "best-worst" answer items were constructed for a later exam. Then the free-answer form could be given in the morning and the objective version in the afternoon, and the two forms correlated to show the faculty what the agreement was between free-answer and objective versions of the same question. The agreement was very high between the free answer and the "best-worst" answer and served to demonstrate to the faculty that objective items did not damage the evaluating power of the exam.

At least some of the free-answer items were graded independently by two persons, usually on a five-point scale, (A,B,C,D, and E), the results plotted and shown to the faculty. The disagreement between the two readers usually astonished the faculty and helped remove the objections to using objective items.

The "best-worst answer" item type is not presented in Sybil Carlson's handbook; only "best-answer" items were included. Since "best-worst answer" gives twice as many responses, it is very probable that its reliability and validity would be higher. Also it requires the student to make judgments similar to those the faculty makes in judging the answers to the parallel free-response item. I feel fairly certain that faculty would resist grading on a two-point scale, (A) vs (BCDE), or (AB) vs (CDE). Grading the items on a

three-or five-point scale requires the student to make judgments similar to those made by the faculty in grading items and demonstrates more clearly that objective items require judgments essentially the same as for essay items.

Turnbull tells me that he has had a study made of "worst-answer" items. He finds that while they do not differentiate among the superior students, they do discriminate among the poorer students, and so they would be a valuable addition to any exam.

Dick Thornton tells me that when he suggested various types of performance items to a home economics teacher (such as having food items cooked and tasted), he was told that a visitor a few years before had told the teacher that only written five-alternative multiple-choice items should be used.

The textbooks of the 60's and 70's also state that writing ten five-choice multiple-choice items in an eight-hour day is good productivity. This slow pace is very likely what discourages teachers from using objective items for their weekly or monthly course exams. Preparing and grading essay exams would be less time consuming than preparing multiple-choice exams if it takes 8 hours to write 10 items.

Using the various non-multiple-choice objective item types indicated above, as illustrated in Appendix A, as well as multiple-choice items where suitable, as in arithmetic and mathematics problems, I found that for classes of ten or fifteen persons, preparing and grading an essay exam was considerably more time consuming than preparing and grading an objective exam that was not primarily multiple-choice.

In discussing exams, a number of teachers have said to me: "I can write a stem and a correct completion, that is no problem, but four plausible false completions are extremely difficult and time consuming."

I tell them to give a free-answer exam first and use the students' wrong answers for the distractors. I should mention here that some of the texts say that having the students prepare the false completions is impractical. I am very puzzled as to where this statement came from. As I have mentioned, we did this routinely at Chicago and found it very convenient and time-saving.

I also tell the teachers that they should try various item types. Simply begin by asking "What do I expect the students to differentiate after having my course that they could not differentiate before?

One answer might be, I expect them to know the difference between the views of John S. Mill, Adam Smith, Kenneth Galbraith, Karl Marx, Frederick Engels, and Lenin. These names would constitute the master list, which would be followed by a series of statements, every one of which would be easy to write for anyone who knows the subject, because it would be a true statement. The student's problem is to assign the statement to the person or persons whose view it expresses. The teacher need not face the problem of devising plausible false statements. In this case, all the statements are true.

Of course, if plausible false statements occur to the teacher, there is no reason to reject them. Simply include them in the list, and add "none of the above" to the master list of choices. It is necessary to specify in the directions that "Only one name is correct for each statement" or that "A statement may express the view of several of these persons."

In an economics exam, one problem would be to distinguish between statements that indicate an increase, a decrease, or no change in supply conditions, or an increase, or a decrease, or no change in demand conditions (see first item in Appendix A). Again the statements are all true. In the psychology exam various abnormalities can be listed, such as Oedipus Complex, Electra Complex, Narcissism, Manic-depressive, Schizophrenia, and others. In this case popular songs are a good source of the symptoms, as for instance:

_____ I want a girl just like the girl that married dear old Dad.

_____ I'm drifting back to dreamland, for I know I'll find you there.

_____ I go wild simply wild over me.

etc.

A similar approach would apply to symptoms of various diseases, to characteristics of various cultures or periods of history, to statements about plays, novels, authors, etc., etc. Often the matrix format is useful for such items because it provides for multiple answers and for omission of ambiguities. The matrix format and many others are illustrated in Sybil Carlson's handbook.

For items of this type, anyone who knows the field can write twenty items in an hour -- not in two eight-hour days. The teacher would find that preparing and grading such items would take far less time than preparing and grading a set of essay items over the same field for a class of about ten or fifteen students. For smaller classes, preparing and grading an essay test might be more economical of time.

One of the item types used at Chicago was "Comment on the Following statement." This was followed by a short quotation from some public figure that was either a good statement that the student should endorse or some ill-advised comment that the student who had learned the points of view developed in the course would object to and could give reasons for the objection. Another similar type of item is "Give a brief answer to the following question." Such items were presented three to a page, with about eight lines allowed for each answer.

The faculty pointed out that grading these items was a difficult task requiring a high order of judgment and understanding of the subject. The grading could be done only by faculty, or possibly by an unusually competent graduate student. A good many of the items were graded twice, independently. The faculty was presented with the scattergraph showing a reasonable correlation of about .8. As they saw it, this constituted astonishing disagreement between the two ratings of the same answer. The answers were saved and used to construct an objective set of items for the next exam. Incorrect answers given by a number of students would be used as false plausible distractors. Requiring the students to grade on a five-point scale, as the faculty had done, turned out to be impractical; so we settled for a set of three answers to each question, or three comments for each statement, the student being instructed to mark "plus" for the best of the three, and "Zero" for the worst of the three and leave the other blank.

During World War II, Norman Frederiksen and I, and about ten other psychologists were on a project organized by John Stalnaker of the

College Board with the Navy to develop aptitude and achievement tests for the Navy schools.

The three years spent on this project were, I think, the most interesting of the sixty-plus years that I have spent in different jobs in the field of psychology. Our work showed that aptitude tests are very useful in assessing the content validity of achievement tests.

The Navy gave a battery of aptitude tests to recruits, including arithmetic, reading, mechanical knowledge, mechanical aptitude, spatial ability, etc. One of the initial findings was that for the gunners' mate schools, the highest correlation was between grades and the reading test. The officers in Washington said, "So we will use the reading test to select recruits for the gunners' mate schools." I said, "Gunners' mates diagnose malfunctioning in guns, and they repair guns. Please let me go to some gunners' mate schools to find out what is going wrong." Norman Frederiksen and I went to Bainbridge Gunners' mate School, and found that grades were based primarily on a final hour-and-a-half written test on the manual. We studied the manual, and it was clear that good reading ability was required to understand it. A practical test was also given to see if the men knew the names and functions of the various parts of a disassembled gun. For this test the chief had a number of gun parts jumbled on his desk, and when the class came in for the examination, he said, "At ease, break out your cigarettes." The class sat down on the floor around the room, smoking and talking. The chief called one man up and picked up a gun part from the array on his desk. "What is this called? What does it do?" A second gun part was used in a similar

manner, and the man went back and was graded on the basis of these two answers. Another man was then called up, asked two questions, and graded on his answers. In an hour or so, each man had answered two questions, and that was the practical exam. I asked the chiefs about the importance of knowing the names, and they were emphatic about its importance. They pointed out that if a part malfunctioned, a new one would have to be obtained. This required the use of the catalogue of parts, and using the catalogue was impossible unless you knew the name.

We then got a large number of pieces of cardboard, wrote down half a dozen part names and functions on each one, distributed about fifty of these cardboards along tables around the room, and on each cardboard we put a piece of equipment whose name and function were given on that cardboard. The students came in and stationed themselves around the room, one at each station. They picked up the piece in front of them, found its name and function on the cardboard, and wrote the appropriate letters or numbers on the answer sheet. At the end of a minute and a half, the signal to move to the next place was given and everyone moved. By this procedure, at the end of 75 to 80 minutes, each student would have answered 100 questions instead of two or four. We observed that for some students the time allowed was ample, and they spent a large part of the minute and a half studying what answer the student just ahead of them was giving to the next question. So for the next exam, instead of saying, "Move to the next position," the proctor said, "Move to the position beyond the next one." Moving two spaces effectively thwarted studying the next item and imposed only the restriction that an odd

number of items be used; so that in making two circuits of the room, each student would have answered all the items.

We also introduced a practical manipulation test where there were some five or ten stations at which the students could disassemble and reassemble a breech block at one station, cock and uncock a gun at another, prepare a gun barrel to be removed from the gun at another, and so on. This latter one was interesting. The chiefs said that removing the gun barrel could not be a test item since it took four men to slide the barrel out because it was so heavy. On learning just what was involved, we saw that the barrel had to be prepared for removal before it could be slid out, and the manipulations necessary for this could be performed by a one-armed soldier. So instead of "Remove the barrel" as an item, we used "Prepare the gun so that a crew of four can slide the barrel out." An instructor was placed at each station to tell the student what tasks to perform and to grade his performance.

These three exams--written, performance, and identification--were scheduled for a three- or four-hour session. During the first half of the session, one-half of the students took the identification test, and the other half the written test. Then for the last half, they reversed. The students who had taken the identification test changed rooms and took the written test, and those who had taken the written test at first now took the identification test. For the performance test, which took only ten minutes, a small group put down their written papers and went out to the performance test room. When they came back, another group went out, so this involved only one interruption of the written test for each student.

When the new grades were assigned on the basis of the tests I have just described, the validity of the reading test dropped, and the validity of the two mechanical tests in the aptitude battery went up.

In one school after another we found a similar pattern. An aptitude test that should have had a high validity had a low one, and another test that should have had a low validity had a high one. Investigating the testing and grading procedures showed that important aspects of the work were inadequately tested and had only a trivial influence on grades, but less important parts turned out to be easy to test and had become very important in determining grades. When we introduced testing procedures that adequately measured what the instructors said were the important aspects of the course, the pattern of validity coefficients changed.

We also found that the exams had a marked effect on the attitude of the students. Before the identification and performance tests were used, much of the students' time in the lab period was spent in smoking and talking with other students. After the first test, the students spent a great deal of time in class at such tasks as assembling and disassembling a piece of equipment. The student would ask another student to time him, and they would compete to see who could get the job completed quicker.

The correlation between aptitude tests and course grades can be used as a way to explore the nature of course evaluation and grading in any school, of course--not just in the military.

For example, in one report of The Psychological Corporation, grades in Latin for one school were best predicted by the clerical tests. This could have been called to the attention of the school to see if

changed methods of teaching, testing, and grading Latin would lower the validity of the clerical test and result in better learning. In some schools mathematics grades are best predicted by the verbal score. In some schools, grades in mechanical drawing have a high correlation with spatial tests, and in others this correlation is very low. What the students are really learning in these two cases must be very different.

It is especially interesting to find that aptitude tests can indicate whether course grades are measuring relevant or unimportant aspects of the course. I had not realized before that aptitude tests could be useful in evaluating the teaching, testing, and grading for a course.

Testing organizations should evaluate the criteria that are available for determining the validity of their tests and, where appropriate, point out that these criteria may well be in need of revision and improvement. Probably, the testing organization should guide the revision of the criteria. It would seem that this overhaul is not undertaken as often as it should be. Teacher-made tests should also be evaluated in terms of the objectives and the desired outcomes of instruction.

I should also mention my experience with the precept system at Princeton University. Small groups of ten or so (the original number when Woodrow Wilson started the system was to be seven students to a group) meet to discuss the lectures of the previous week. The system was designed as a gathering in which students could raise questions and comment on the lecture and study material. I would come to the class and open it by saying, "Are there any questions or comments on the material you have studied?" I was usually met with silence. Then I would ask

questions or make some review comments, in an attempt to get a discussion started. But when I finally prepared a five- or ten-minute quiz on the material and started the class by giving the quiz, having papers swapped with a neighbor for scoring the quiz, the comments and questions came immediately and in abundance, and needed no prompting from me.

I'll summarize by stressing again a few points that seem most important to me.

Teachers should learn that the multiple-choice item type is to be used only when it is an appropriate item form. For example, in arithmetic and mathematics, multiple-choice items should have for distractors the errors frequently made by the students. However, in arithmetic and mathematics, the free-answer item form is very convenient and very easy to score reliably, since there is a definite and unambiguous correct answer in these fields. When machine scoring is not used, the free-answer form is scored as quickly as the objective form for math and arithmetic.

Generally speaking, the teacher should learn how to use the item form that is most natural and convenient for the particular subject at hand. The "master list" is frequently the most convenient and suitable item type. However, in some situations it is most appropriate to use other formats such as "greater-less-same," "matrix," "best-worst answer," and other item types, as described in Sybil Carlson's Creative Classroom Testing: Ten Designs for Assessment and Instruction (ETS, 1985).

I also think we might well re-examine the possible usefulnes of item types other than multiple-choice in tests prepared by ETS.

Another point that should be stressed in training teachers to write items, is the usefulness of performance and identification items, especially in courses that include some laboratory work.

Also, the improved design of free-answer items should be included, as illustrated by the work of Ward, Frederiksen, and Sybil Carlson. I refer here to medical diagnosis items, in-basket items, and others of the free-answer item type.

These various types of items would improve the accuracy of teacher testing and grading and also facilitate and promote student learning.

The problem is, what steps can be taken and by whom, to teach teachers to write easy and appropriate items?

When teachers come to ETS for a session on item writing, various item types should be stressed and illustrated. Handbooks like Carlson's should be prepared on how to construct and grade various sorts of free-answer items. Performance and identification items, incidentally, have been found to be very useful in the work of Richard Thornton, and Michael Rosenfeld.

In addition to ETS's special seminars dealing with various item types, work of this sort should be made a standard part of teacher training. Ideas about methods effective for getting such training under way in teacher-training schools are needed, making Sybil Carlson's handbook "Creative Classroom Testing" and additional handbooks on "Performance Items" and various types of "Free Answer Items" readily available to teachers might be an effective way of getting such training underway.

Annotated References

Adkins, D. C. (1974). Test construction, development and interpretation of achievement tests (2nd ed.). Columbus, OH: Charles E. Merrill. (Deals with essay and various types of objective tests; a final chapter on criticisms of tests.)

Adkins, D. C., Primoff, E. S., McAdoo, H. L., Bridges, C. F., & Forer, B. (1947). Construction and analysis of achievement tests: The development of written and performance tests of achievement for predicting job performance of public personnel. Washington, DC: United States Government Printing Office. (Gives various types of objective items, and a final chapter on Performance Tests.)

Ahmann, J. S. (1962). Testing student achievements and aptitudes. Washington, DC: Center for Applied Research in Education. (Supply type, Multiple choice, Matching, Essay)

Beggs, D. L., & Lewis, E. L. (1975). Measurement and evaluation in the schools. New York: Houghton Mifflin Company. (See especially: Pp 158-165, Multiple-choice, completion, short answer, matching, Essay; Pp 174-181, Multiple-choice, True-false, matching, completion, short answer, Essay.)

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill. (Numerous illustrations)

Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). Evaluation to improve learning. New York: McGraw-Hill.

Burt, C. (1948). Handbook of tests for use in schools (2nd ed.). London: Staples.

Carlson, S. B. Creative classroom testing: Ten designs for assessment and instruction. Princeton, NJ: Educational Testing Service.

Chauncey, H., & Dobbin, J. E. (1963). Testing: It's place in education today. New York: Harper and Row.

Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods: New York: Irvington.

Ebel, R. L. (1965). Measuring educational achievement. Englewood Cliffs, NJ: Prentice-Hall. (See especially: Ch. 4 Essay Tests, Ch. 5 True-false; Ch. 6 Multiple-choice.)

Furst, E. J. (1958). Constructing evaluation instruments. New York: Longmans, Green. (See especially: Ch. 9 Supply-type items, extended answer, short answer; Ch. 10 Choice type items, Multiple-choice, greater realiability per item; Classification - Key-list, or master list.)

Hawkes, H. F., Lindquist, E. F., & Mann, C. R. (1936). The construction and use of achievement examinations. New York: Houghton Mifflin. (Most of the book illustrates test items.)

Monroe, W. S., DeVoss, J. C., & Kelly, F. J. (1924). Educational tests and measurements. New York: Houghton Mifflin.

Richardson, M. W., Russell, J. T., Stalnaker, J. M., & Thurstone, L. L. (1933). Manual of examination methods. Chicago, IL: University of Chicago, Board of Examinations.

Rinsland, H. D. (1937). Constructing tests and grading in elementary and high school subjects. New York: Prentice-Hall.

Roid, G. H., & Haladyna, T. M. (1982). A technology for test-item writing. New York: Academic Press.

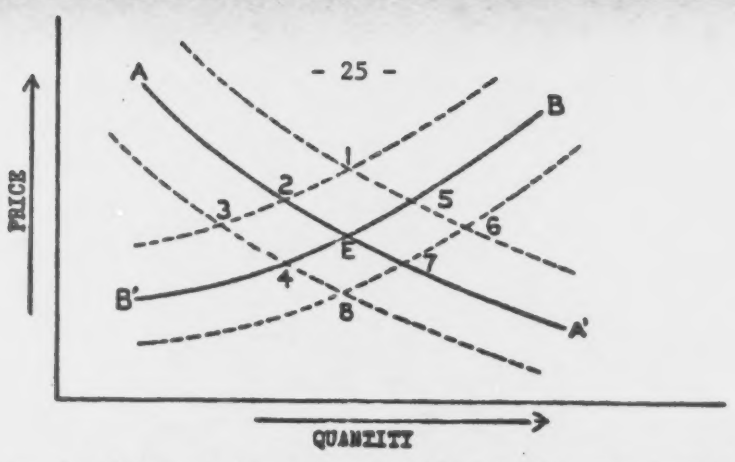
Rosenfeld, M., & Thornton, R. F. (1978). The development and validation of a firefighter physical selection test. Princeton, NJ: Educational Testing Service.

Ruch, G. M. (1929). The objective or new-type examination: An introduction to educational measurement. New York: Scott Foresman. (See especially: Ch. VIII Pp 188-212.)

Tyler, R. W. (1934). Constructing achievement tests: Reprints from the Educational Research Bulletin. Columbus, OH: The Ohio State University Press.

Appendix A

Sample Test Items



Assume that, in the diagram given above, AA' and BB' represent the original supply and demand conditions for the product indicated in column B. The equilibrium price is represented by the point marked E.

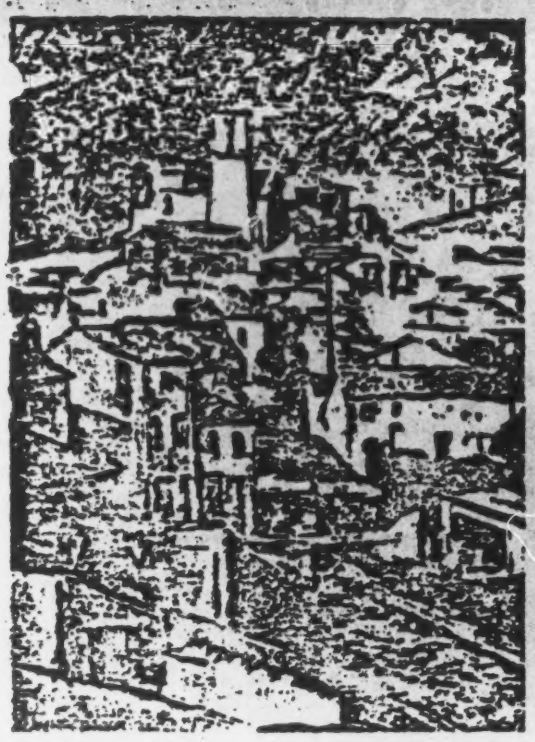
Each of the following statements represents a hypothetical change in these original conditions.

The effect of these changes in producing a new equilibrium price can be illustrated by one of the numbered equilibrium points shown in the diagram.

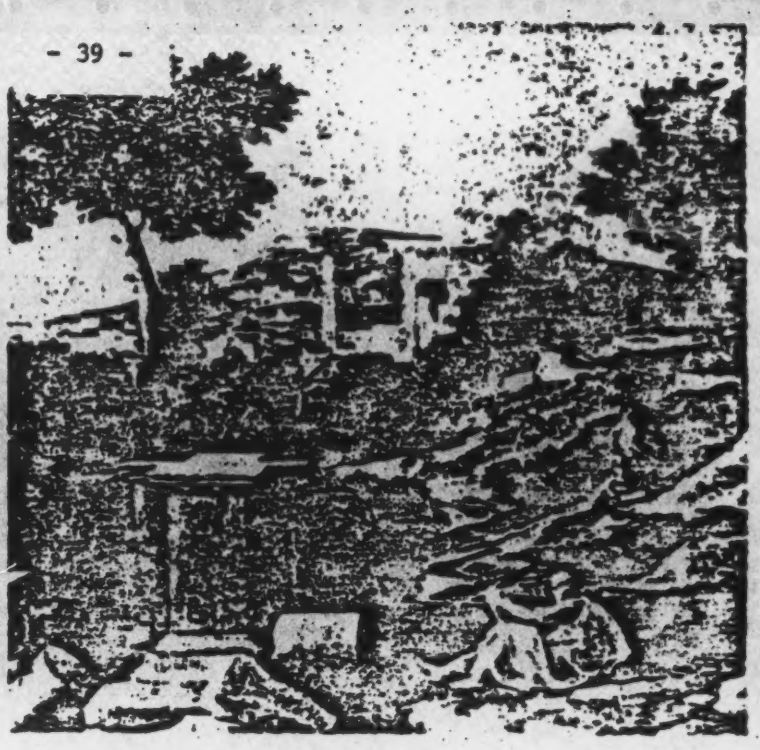
In each blank space in column A below, write the one number which best illustrates the location of the new equilibrium point (for the commodity indicated in column B) which will result from the change in supply and/or demand conditions indicated.

(In each case assume that no other influences are operating.)

Column A	Column B	Statements
_____	Silk Hosiery	The increased preference for silk to cotton hosiery has been met by the introduction of automatic silk knitting machines.
_____	Steel	The World War increased the strain on the productive capacity of the steel industry, at the same time that the draft act took a large number of laborers out of the steel industry.
_____	Gasoline	There has been an increase in the number of automobiles in use, as well as in the yearly mileage per automobile (number of miles traveled per year by each automobile).
_____	Phonographs	Phonographs can be manufactured just as efficiently as ever, but people do not seem to care for them any more.
_____	Wheat	The drought has reinforced the effect of the wheat reduction program and at the same time unemployment has slashed the purchasing power of wheat users.
_____	Carriages	Machine production has displaced hand production in the carriage industry, and at the same time the carriage has been displaced by the automobile.
_____	Cottonseed	There has been an increase in the demand for cotton, but the demand for cottonseed has remained the same.
_____	Lumber	Since virgin forests are being depopulated, the lumber industry is becoming more dependent on second growth timber.



Picture K



Picture L

Key for items 51 - 60
1. Applies to picture K.
2. Applies to picture L.
3. Applies to both of these.
4. Applies to neither.

51. The picture is basically an organization of lines and geometric forms.

52. The artist evidently believes that a painting should have structural coherence.

53. A notable characteristic of the painting is the shimmering insubstantiality of the forms represented.

54. The artist presents a version of nature in which logic and order prevail.

55. A notable characteristic of the picture is loose informality of its structure.

56. The picture is a study in the regression of successive planes.

57. In comparison with most of the painting of the nineteenth century, the picture has a relatively flat schematic plan.

58. The style strongly suggests the work of Monet.

59. The style suggests strongly that the picture is the work of Cezanne.

60. The picture is a Cubistic work by Picasso.

Time: 7 Minutes

Write the appropriate number or numbers in each space.

PORTAL OF ENTRY	CASUAL ORGANISM	AGENT OR MEANS OF TRANSMISSION	METHOD OF CONTROL
1. skin wound	1. virus	1. mosquito	1. screening homes
2. respiratory tract	2. bacterium	2. flea	2. killing rats
3. urogenital tract	3. protozoa	3. dog	3. vaccination or inoculation
4. intestinal tract	4. none of these	4. man	4. quinine
0. none of these		5. water	5. antitoxin
		6. cow	6. draining swamps
		7. milk	7. pasteurise milk
		0. none of these	8. purifying water
			0. none of these

DISEASE	PORTAL OF ENTRY	CAUSAL ORGANISM	AGENT OR MEANS OF TRANSMISSION	METHOD OF CONTROL
Malaria				
Smallpox				
Syphilis				
Typhoid				
Rabies				
Diphtheria				
Beri-beri				

Matrix

Topic: Child Development

Content Objective: Differences in thinking in Piaget's major stages

Behavioral Objectives: Knowledge of classifications, principles and generalizations, theories and structures; comprehension (translation and interpretation) and analysis of elements and relationships

* * *

Piaget's Stages of Thinking

Place an X in the appropriate squares to indicate the forms of thinking that are characteristic of, or make their appearance during, each stage of mental development. (It is possible to check the squares for more than one stage.)

	I Sensorimotor	II Preoperational	III Concrete Operations	IV Formal Operations
1. focus on states		X		
2. formulates hypotheses				X
3. egocentrism	X	X	X	X
4. irreversibility		X		
5. decentering			X	
6. object permanence	X			
7. true imitation	X			
8. the symbolic function		X		
9. accommodation	X	X	X	X
10. memory	X			
11. deferred imitation	X			
12. transformations of state			X	
13. operations on operations				X
14. transductive reasoning		X		
15. conservation			X	

3. Tabular (Matrix)

Topic: General Science--the solar system

Content objective: Information about the different planets--combined

Behavioral objective: Knowledge of specific facts and analysis of elements

* * * * *

The Planets of our Solar System

Fill in the blanks by choosing from the answers provided.

Position		Size	No. of satellites		God after which planet is named	
1. 1st	6. 6th	A. larger earth	A. C	E. 9	A. war	E. seed sowing
2. 2nd	7. 7th	B. smaller earth	B. 1	F. 12	B. love	F. underworld
3. 3th	8. 8th		C. 2	G. none of these	C. winged god	G. none of these
4. 4th	9. 9th		D. 3		D. sea god	
5. 5th						

Planet	Position from Sun	Size Relative to Earth	No. of Known Satellites	God after which Planet Named
Earth	3	X	B	X
Jupiter	5	A	F	G
Mercury	1	B	A	C
Uranus	7	A	G	G
Saturn	6	A	E	E
Pluto	9	B	A	F
Mars	4	B	C	A
Venus	2	B	A	B
Neptune	8	A	C	D

Answer the following questions.

Race A attained a high level of civilization three thousand years before race B attained a similar level of civilization. What is the best interpretation of this fact?

What is the relation of hereditary mental traits of a race to the progress of that race in civilization?

If you were the leader of a group that wished to perpetuate mores sharply divergent from those in western European culture, what measures would offer you the greatest likelihood of success? Ignore considerations of the practical feasibility of the measures, i.e., assume that the measures could be put into effect and outline a plan which would prevent the group from adopting western European mores.

Below are a number of questions.
There are three answers to each question.

Write a plus (+) before the one best answer in each group.
a zero (0) before the one poorest answer in each group.

Question: Race A attained a high level of civilization three thousand years before race B attained a similar level of civilization. What is the best interpretation of this fact?

Answers:

_____The difference in time is so slight in comparison to racial history that it is probably a chance difference.

_____Race B probably had environmental handicaps which did not apply to race A.

_____The probability is that race A is innately better equipped than race B.

Question: What is the relation of hereditary mental traits of a race to the progress of that race in civilization?

Answers:

_____There is hardly any evidence to show that the anatomical characteristics of the races which possess the highest civilization are different from those of races which are on lower levels of culture; but there may be some justification for the belief that hereditary mental traits are to some extent correlated with culture level.

_____Since there is no good criterion of cultural advancement of a race, and we know little aout the relationship between mental and anatomical traits, the most probable view is that there is no necessary connection among the three variables, hereditary mental traits, anatomical traits, and culture level.

_____Actually there are observable correlations between hereditary mental traits, anatomical traits, and culture level, and a causal relation exists between these things.

Question: If you were the leader of a group that wished to perpetuate mores sharply divergent from those in western European culture, what measures would offer you the greatest likelihood of success? (Assume in each case that the measure is practically feasible and can actually be put into effect).

Answers:

_____Frame your program in terms that would be attractive to existing dissatisfied groups, gain publicity and members of the movement, organize a movement and develop local leaders, gain more and more political power until you could seize control of the state, and then consolidate your position and enforce the mores on the members of the group, making wealth, prestige, and power the rewards of conformity.

_____Establish a separate school system in which you would train the children in your group, and those of as many converts as you could get, in the belief in the mores of your group. Emphasize the use of rituals and ceremonies to develop strong emotional reactions around the symbols and slogans of your group in the children at a very early age and on regular and frequent occasions thereafter. Foster in all the members of the group a belief that they are the bearers of a very special responsibility to preserve and spread to the rest of the world the sacred mores of your group.

_____Take your group to a naturally isolated location and establish a separate, self-sufficient community, completely out of touch with the rest of the world. Develop your own institutions and traditions. Carefully guard against travel by the members of your colony and visits from outsiders.

Statement and comment

Topic: U.S. History

Content Objective: the understanding of mercantilism as learned from the text and from classwork, by interpreting a quotation from Adam Smith's Wealth of Nations (not seen previously).

Behavioral objectives:

* * *

"To prohibit a great people, however, from making all that they can of every part of their own produce, or from employing their stock and industry in the way that they judge most advantageous to themselves, is a manifest violation of the most sacred rights of mankind. Unjust, however, as such prohibitions may be, they have not hitherto been very hurtful to the colonies. Land is still so cheap, and consequently, labour so dear among them, that they can import from the mother country, almost all the more refined or more advanced manufactures cheaper than they could make them for themselves. Though they had not, therefore been prohibited from establishing such manufactures, yet in their present state of improvement a regard to their own interest would, probably, have prevented them from doing so. In their present state of improvement, those prohibitions, perhaps, without cramping their industry, or restraining it from any employment to which it would have gone of its own accord, are only impertinent badges of slavery imposed upon them, without any sufficient reason, by the groundless jealousy of the merchants and manufacturers of the mother country. In a more advanced state they might be really oppressive and insupportable."

Pick the best (+) and worst (-) interpretation of this statement from each group of three comments below. Base your answers on the statement itself and your overall knowledge of history.

1. ____a. The statement describes France and her American colonies.
   __-__b. The statement describes Spain and her American colonies.
   __+__c. The statement describes England and her American colonies.

2. ____a. The author believes a mother country can regulate the business of a colony without necessarily hurting the colony.
   __+__b. The author believes people should be left free to develop to their greatest economic potential.
   __-__c. The author believes mother countries should be careful only to prohibit those industries in their colonies that would not develop anyway.

3. __+__a. The author believes the mother country placed the manufacturing prohibitions on its colony because of unnecessary jealousy.
   ____b. The author believes the mother colony placed the manufacturing prohibitions on its colony as a sign of domination.
   __-__c. The author believes the mother country placed the manufacturing prohibitions on its colony to save the colony from higher labor costs.

4. ___ a. The author is arguing against the economic theory of capitalism.
   + b. The author is arguing against the economic theory of mercantilism.
   - c. The author is arguing against the economic theory of Laissez Faire.

5. + a. This statement was most likely made by Adam Smith.
   ___ b. This statement was most likely made by Thomas Jefferson.
   - c. This statement was most likely made by Charles Townshend.

Below are a number of questions for which you will find three different answers. Write a plus (+) before the best answer in each group, and a zero (0) in front of the poorest answer in each group. One answer will be left blank for each question.

Question: Rabbits who live in regions of heavy snowfall all have white fur. What is the best interpretation of this fact?

Answers:

0 The rabbits acquired the trait from their environment and passed it to their offspring.

_____ A mutation occurred in normal rabbits producing white fur.

\+ A mutation occurred in dark furred rabbits producing offspring with white fur and a better chance of surviving in a white environment.

- - -

Question: Spontaneous generation is not generally accepted today, even though it probably has occurred before. The best explanation for this is:

Answers:

_____ We cannot explain how life began.

0 Some types of organisms can only be reproduced in the laboratory.

\+ At one time life did not exist on earth, so life had to begin in some way other than the process of biological reproduction.

- - -

Question: The steady state theory of the creation of the universe has been weakened considerably due to the discovery of:

Answers:

0 The disappearance of fossils at a level corresponding to about 4.5 billion years ago.

_____ The red shift of light at the edge of our galaxy indicating our galaxy is expanding.

\+ The red shift of light from other galaxies indicating they are moving away at tremendous speeds.

There are three comments on each of the following statements. Mark
plus (+) for the one best comment on each statement;
zero (0) for the one worst comment on each statement.

Statement: I do not think that forcing a naturally left-handed child to use his right hand for writing and eating is of any particular importance in causing stuttering. Studies which have been made are in agreement in showing that 90 to 95 per cent of the children who are made to use the right hand instead of the left never stutter.

Comments:

_____ The person making the foregoing statement shows his lack of knowledge of the literature; some studies do show 90 to 95 per cent who do not stutter, but the majority of better studies show this percentage to be about 50 or 60. In other words we cannot say that a forced shift of hand preference is the major factor in causing stuttering, but we can say that it is one very important factor.

_____ There is a major fallacy involved in this line of reasoning. The percentage of stutterers who were forced to use the unpreferred hand would be a much better index of the possible relevance of this factor to stuttering than would the figure which is given.

_____ The argument is correct as far as it goes, but it is necessary to recognize that multiple causation may be important in the etiology of stuttering and hence other possible causes of stuttering should be investigated in the same way and comparisons made before reaching any conclusions.

Statement: Emotions are primarily physiological disturbances and are best differentiated in terms of the corresponding physiological changes.

Comments:

_____ The statement is essentially correct. Physiological changes show specific patterns corresponding to each of the important emotions.

_____ As McDougall has so well demonstrated, introspective reports of subjects regarding their emotions must also be considered in classifying the emotions.

_____ The statement is definitely misleading--for example, it is impossible to distinguish between fear, rage, and excitement on the basis of the physiological changes.

Statement: Most human behavior is directed by unconscious motives.

Comments:

_____ It is true that people are frequently unaware of the importance, and sometimes even of the existence, of many factors which influence their behavior.

_____ There is no such thing as unconscious behavior.

_____ So-called unconscious processes are better explained in terms of blind trial and error.

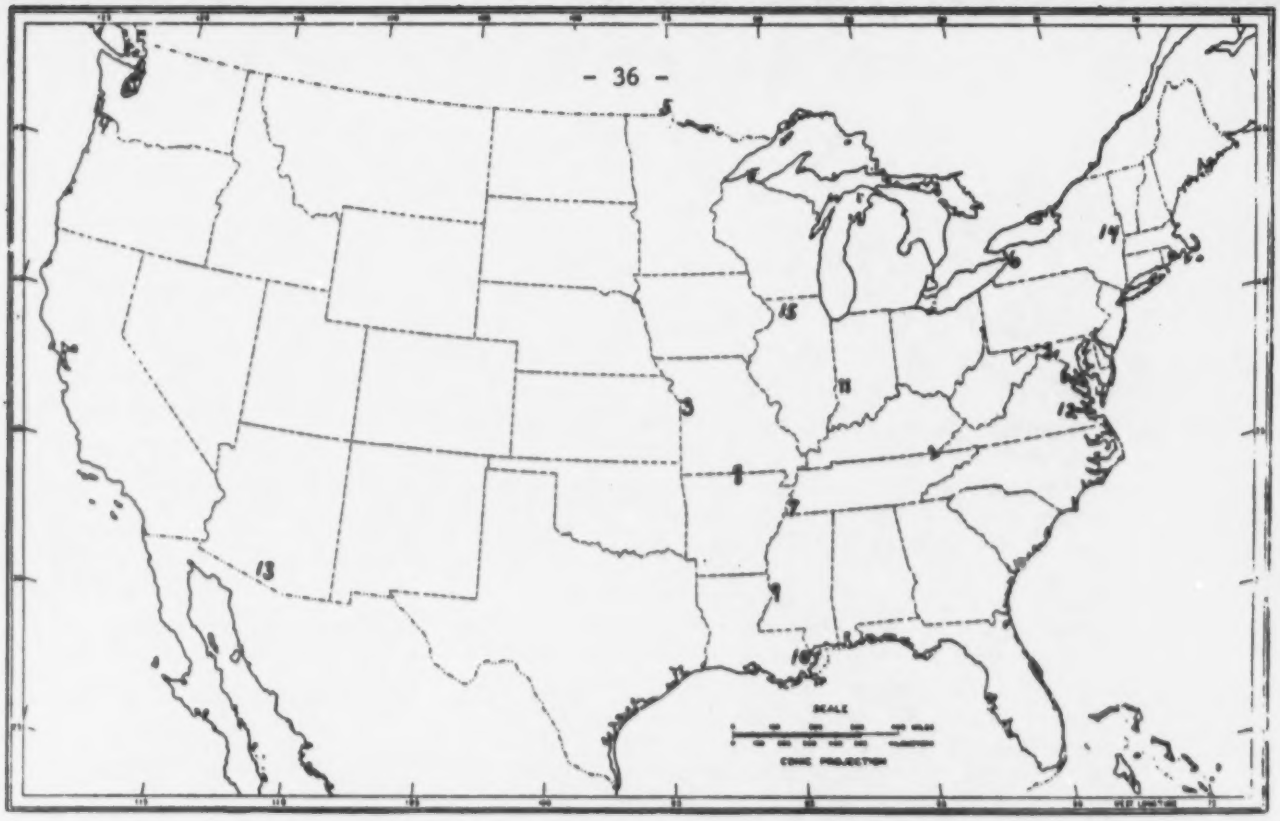
Statement: A man's interests and desires may become ever so elaborate, refined, socialized, sublimated, idealistic; but the raw basis from which they are developed is found in the phenomena of living matter.

Comments:

_____ Many drives have definite physiological bases.

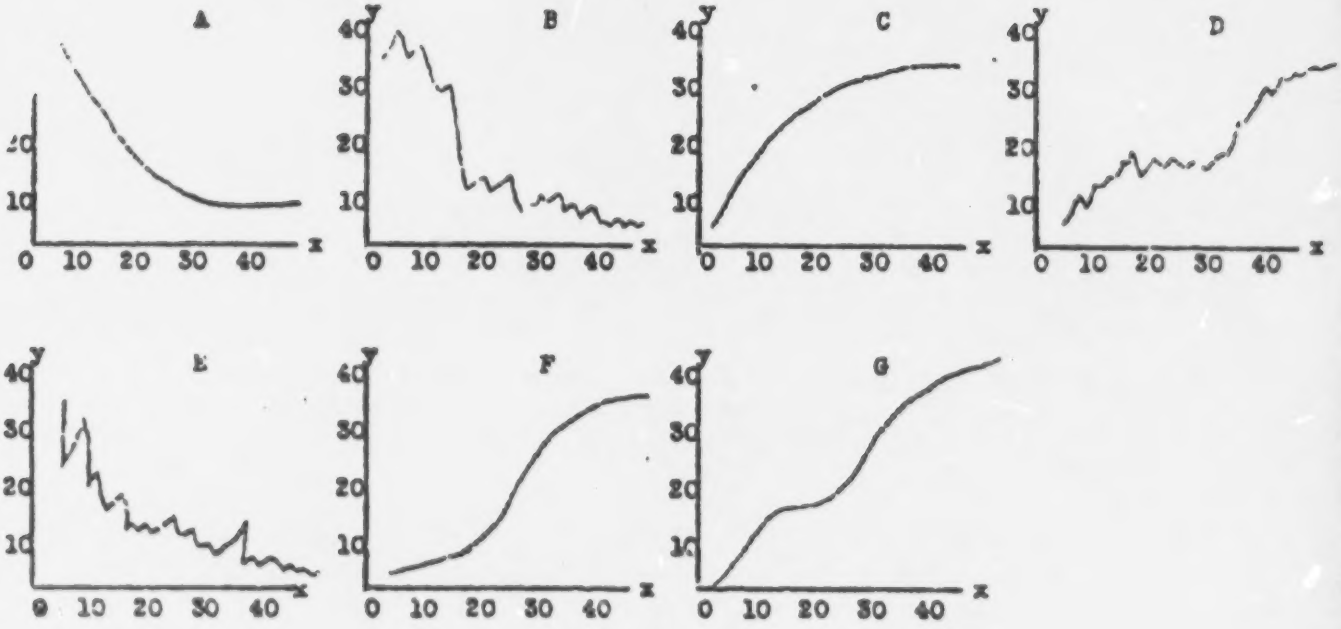
_____ The social drives are independent of physiological factors.

_____ The social drives develop out of the satisfaction of physiological drives by other people.



Each of the following places or locations is represented by a number on the map at the top of the page. In addition each of the names may be associated with one of the descriptive phrases at the right. There are more numbers on the map and more descriptive phrases than names in the list. In the first blank before each name, place its map number. In the second blank, write the letter of the description which best fits it.

Map Number	Letter		
15	H	Freeport	A-eastern end of Santa Fe-Oregon trails
13	G	Gadsden Purchase	B-early American fur-trading post in the Far West
12	I	Richmond	C-The Wilderness Road went here.
	D	Harpers' Ferry	D-John Brown raided it.
3	A	Independence	E-strategic location for control of Mississippi River in the Civil War
6	J	Buffalo	F-Missouri Compromise line
4	B	Astoria	G-connected with the question of a transcontinental railroad in the 1850's
9	E	Vicksburg	H-Douglas tried here to reconcile popular sovereignty and the Dred Scott opinion
1	F	36° 30' latitude	I-It took almost four years for the North to capture it.
11		1940 center of United States population	J-eastern end of Erie Canal
2	C	Cumberland Gap	



Encircle the letters representing the curves (or curve) to which the following description apply. Ignore the alternatives which are crossed out.

1. The beginning of the curve shows positive acceleration. ~~A~~ B C D ~~E~~ (F) ~~G~~
2. The ordinate values (y-values) could denote errors per trial. (A) (B) C D (E) F G
3. Points on the curve probably represent average values obtained from more than two or three subjects. (A) B (C) D E (F) (G)
4. Curve might be one which represents the results from a task in which the learner displayed "insight". A (B) C D E F G
5. This curve could be a sample Thorndike curve showing the rate of learning of one dog or one cat in a problem box. A (B) C ~~D~~ (E) F G
6. The ordinate values (y-values) could denote time per trial. (A) (B) C D (E) F G
7. This is probably the learning record of one individual. A (B) C (D) (E) F G
8. The ordinate values (y-values) may denote cumulative errors. A B (C) D E (F) (G)
9. The curve contains a plateau. A B C (D) E F (G)
10. The ordinate values (y-values) may be the number of correct responses per trial. A B (C) (D) E (F) (G)
11. The abscissa values (x-values) may be cumulative trials or cumulative time. (A) (B) (C) (D) (E) (F) (G)

Rank Order

Topic: English composition

Content objective: Coherence of an explanatory paragraph

Behavioral objectives: Knowledge of conventions, comprehension (interpretation), application to a new situation, and synthesis

* * *

Writing a Logical Paragraph

The sentences given are in proper form to make a coherent paragraph, but they are not in the proper sequence. Arrange them logically, placing the appropriate number next to each sentence.

2 A suspension bridge is based on the theory that if cables can be strung across an area, a bridge can be hung from them.

4 The roadway platform is supported by vertical cables, which in turn are attached to the dipping cables.

1 The most practical kind of bridge for use in spanning wide areas is the suspension bridge.

7 This will allow for stabilization of the bridge, and will make it more secure.

3 The bridge is made of sturdy steel cables which are supported by towers and secured in the ground at either end.

5 Girders then support the roadway, in order to prevent the bridge from movement.

6 According to most engineers, the main span of a suspension bridge should not be more than 7,000 feet in length.

The Normal Course of Motor Development (birth to age 3)

The following paired statements refer to milestones in infant motor development that generally are acquired sequentially. If the motor skill described on the left appears before the motor skill described on the right, encircle the letter "B"; if the skill on the left appears after the skill on the right, encircle the letter "A"; if the skills on the right and left appear at approximately the same time, encircle the letter "S".

Skill	When does it appear?	Skill
head control	A (B) S	rolling over
sitting up	(A) B S	crawling
turning from back to stomach	(A) B S	turning from stomach to back

In a study of reasoning Maier trained rats in the following way:

(1) On some days each rat was allowed to explore a table top which had food in corner A, a box in corner B, and a low wall across the middle of the table separating corners A and B. The rats could and did climb over this wall.

(2) On other days each rat was placed in the box in corner B and learned to run a maze leading out from the table and back to corner A.

(3) Rats were placed on the table, but prevented from reaching the food in corner A by a transparent screen.

Indicate your knowledge of the results of this experiment by marking each of the following items:

plus (+) if true,
zero (0) if false.

_____ The rats succeeded in forcing their way through the screen and reached the food.

_____ The rats attempted unsuccessfully at first to get through the screen to the food.

_____ The rats gave up trying to reach the food and rested in a corner of the table.

_____ The rats took the maze pathway to the food, running with essentially no errors.

_____ The rats took the maze pathway to the food but were obviously disturbed, making many errors on the way.

_____ The behavior of the rats was interpreted as indicating a trial and error solution of the problem.

_____ The behavior of the rats was interpreted as indicating frustration and regression.

_____ The behavior of the rats was interpreted as indicating an insight solution.

_____ The rats showed a typical form of neurotic behavior.

Keylist

Topic: U.S. History

Content objective: the Mexican War and Maine and Oregon boundary disputes

Behavioral objectives:

* * *

Indicate the relationship of the items listed below as follows:

C for items that were causes of the Mexican War

E for items that were effects of the Mexican War

O for items that had nothing to do with the Mexican War but had to do with the Oregon boundary dispute

M for items that had nothing to do with the Mexican War but had to do with the Maine boundary dispute

C 1. Election of James K. Polk

M 2. Webster-Ashburton Treaty

E 3. Formation of the Bear Flag Republic

M 4. Battle of the Maps

C 5. Annexation of the Lone Star Republic

M 6. Aroostook War

E 7. Mexican Cession

C 8. Crossing the Nueces River by Gen. Taylor

D 9. American settlers in the Wilamette Valley

C 10. Manifest Destiny

C 11. U.S. desire for California

E 12. Wilmot Proviso

Statement and Comment

Topic: Child Development

Content objective: Integration of the basic concepts in the unit on adolesent development

Behavioral objectives: Knowledge of specific facts and of principles and generalizations, comprehension (interpretation), and application to concrete situations

* * *

For each of the following statements, indicate which comment about the typical adolescent is the most appropriate explanation for the statement.

Comments:

A. The intellectualism of adolescence is seen as an ego-defense mechanism.
B. The intellectualism of adolescence is seen by cognitive theorists as a "flexing of intellectual muscles."
C. Most adolescents are at the conventional stage of morality (level II).
D. The adolescent is continually aware of an "imaginary audience."
E. The adolescent is influenced by the "personal fable."
F. Adolescents undergo identity confusion as they pursue their search for identity.
G. The adolescent is searching for a redefinition of the relationships to parents.
H. Adolescents are highly influenced by their peer group.

Statements:

D 1. Susan spends hours in front of the mirror, putting on makeup and blow-drying her hair, because she desires a nice appearance at all times.

E 2. Steve's parents are concerned about his wild motorcycle driving, but he thinks they are too stuffy.

C 3. Joan strives very hard to do well in English, because her teacher is so wonderful.

A 4. Harold has become deeply involved in a "born again" religious movement.

E 5. Jim feels that his parents are too old-fashioned, and are not sufficiently aware of the pressures on young people today to sign up for the draft.

(this item would be continued)

Properties of Elements

A. alkali metals	E. hydrogen	I. neon
B. francium	F. iodine	J. noble gases
C. halogens	G. fluorine	
D. helium	H. lithium	

G noble gases (choose 2)

D or F

B alkali metals (choose 2)

H

F halogens (choose 2)

G

D the only substance that does not form a solid at any temperature.

G one of the few elements that does not form oxides, hydrides, and fluorides.

E an unusual element because its chemical properties indicate that it could be grouped with either lithium or fluorine.

B a very rare alkali element.

A very high chemical reactivity.

H very low chemical reactivity.

A chemical behavior is characterized by ionic bonds.

C react with hydrogen gas to form hydrogen halides.

H lowest boiling temperatures known.

In the blank space before each of the following statements write the number of the <u>one best</u> term.

1 transference	6 introversion
2 regression	7 extroversion
3 projection	8 ambivalence
4 manifest content	9 conversion
5 latent content	

4 The subject reported a dream of being lost in a snowstorm, and of being very cold, especially on his hands and feet.

2 Soon after he lost his job, Mr. X, who had spoken excellent English for years, could only speak Polish, which was the language of his childhood.

2 Is characterized by withdrawal tendencies and the adoption of habit patterns long since discarded.

1 During the course of the analysis all the resentment which Ruth had felt for her parents was felt toward the psychoanalyist.

9 A bookkeeper who disliked his job suddenly developed a functional blindness.

7 Trust no Future, howe'er pleasant!
Let the dead Past bury its dead!
Act, -- in the living Present!
Heart within, and God o'erhead!

3 The pot calls the kettle black.

5 An analysis of the dream showed that it referred to a basic conflict between the child and its parents.

6 Here where the world is quiet, here where all trouble seems
Dead winds and spent waves riot, in doubtful dream of dreams.

8 Yet each man kills the thing he loves, by each let this be heard
Some do it with a bitter look, some with a flattering word.

A series of four short musical selections will be played for you. The series will be played three times with brief intervals between playings. For items 136-45 choose the appropriate response from the following key list.

1. Applies to the first selection.
2. Applies to the second selection.
3. Applies to the third selection.
4. Applies to the fourth selection.
5. Applies to none of the selections.

136. Is polyphonic.

137. Is distinguished by grace, urbanity, clear and sharply-cadenced phrasing, melodic charm.

138. Is distinguished by its untraditional harmony and by its lack of dynamic contrasts.

139. Is similar to the work of Bach in texture; is untraditional in harmony.

140. Is Classical.

141. Is Impressionistic.

142. Is the work of a twentieth century composer.

143. Might be an excerpt from the third movement of a four-movement symphony by Mozart.

144. Was written before the Classical Period.

145. Illustrates the musical style characteristic of the period between Beethoven and Debussy.

****

146. (Your response to this item will not affect your examination score.) I feel that the time allowed for the items related to musical selections played during the examination has been

1) insufficient 2) about right 3) too generous.

GRAMMAR I: COMPLETION

Suggested time: 30 minutes

A

1. Write out in French words the number 1794._____

In every lettered group of sentences below, each sentence may lack a word or group of words, whose position is indicated by a blank --. The sentence, if incomplete, will be correctly completed when one of the lettered words or phrases listed at the left is inserted in the blank. Decide which of the items will correctly complete the sentence and put its letter in the space at the right. (Do not put it in the blank of the sentence itself.)

Example: a. des	J'ai --- tête.	c
b. du		
c. une	Voyons -- livres.	a
d. de		
e. no addition needed	A-t-il -- café?	b

B

a. prudentment	2. Il agit --.	2._____
b. prudemment		
c. prudentement		

C

a. es	3. Tu -- couru vite.	3._____
b. as		
c. êtes	4. Vous vous -- blessée.	4._____
d. avez		
e. ferez	5. Vous me -- pleurer.	5._____
f. est		
g. devons	6. Nous -- partir.	6._____
h. a		
i. sommes		
j. no addition needed		

D

a. février douze	7. On célèbre la fête de Lincoln --.	7._____
b. février douzième		
c. le douze février		

E

a. me
b. mes
c. la
d. de
e. de la
f. les
g. leurs
h. no addition needed

8. Tout le monde craint — mort. 8._____
9. Il possède une collection — tableaux. 9._____
10. Il doit avoir -- honte. 10._____

F

a. ce
b. cela
c. celui
d. celle
e. ceux
f. ces
g. cet
h. cette
i. no addition needed

11. J'ai ma carte et — de Jean. 11._____
12. Parmi les livres j'ai choisi un seul: — là. 12._____
13. Je comprends -- que vous voulez dire. 13._____

PART III - READING COMPREHENSION

Directions: Each one of the eight selections in this Part is followed by a number of questions based on the reading passage and answerable usually by a single word or a short phrase. Complete sentences are NOT necessary or desired in answering. The questions do not necessarily follow in the order of the development of the passage. After having read the passage through carefully, write in ENGLISH your answers to the questions within the parentheses to the right. Write plainly. Note that the passages increase in difficulty of content as you near the last selections; you should allow more time for the latter half of this section than for the first half.

Time allowed: 30 minutes

French 104-105-106
Reading Comprehension--Form A

I. Une jolie actrice, Mademoiselle Lanlaire, jouant sur le théâtre de Bordeaux, manqua son entrée d'une demi-heure et en véritable enfant gâtée, reçut fort mal les témoignages de mécontentement du public. De là, tapage et sifflets chaque soir à son entrée, de là aussi intervention active des troupes du gouverneur qui avait ses raisons pour protéger la belle. Les sifflets et les cris interdits, on trouva autre chose pour les remplacer. Dès que Mademoiselle Lanlaire paraissait, on se mit à tousser; mais la prison attendait ces malades. Enfin l'un des conspirateurs s'avisa d'apporter au théâtre un jeune chien; aussitôt que l'actrice se montre, il pince la bête, qui remplit la salle de ses cris. Mademoiselle Lanlaire, qui s'était retirée, a le courage de reparaître; un brutal lui jette un soulier à la tête. Cette fois, on ferme les portes du parterre qu'on fait évacuer par une seule issue. Ainsi l'homme au soulier ne pourra échapper. Le premier qui se présente n'est chaussé que d'un pied. "C'est lui!" s'écrie le soldat de droite, qui le voit. Mais le second n'a qu'un soulier non plus. "Le voilà!" fait le soldat de gauche. "Non, c'est celui-ci!" dit un autre soldat en saisissant le troisième. Tout le parterre s'était déchaussé le pied gauche. Il fallait bien laisser passer tranquillement tout le monde.

---

1. De quel caractère l'actrice semblait-elle? (_____)
2. Comment avait-elle déplu aux spectateurs? (_____)
3. Qu'est-ce qui marquait chaque soir plus tard l'arrivée de l'actrice? (_____)
4. De qui était-elle la favorite? (_____)
5. Qu'est-ce que son public a fait lorsqu'on eut interdit les sifflets et les cris? (_____)
6. Plus tard encore, que fait celui qui a un chien quand Mlle. Lanlaire paraît? (_____)
7. Qu'est-ce qu'un brutal lui jette à la tête? (_____)
8. Ce soir-là par quelles portes faut-il que le parterre sorte? (_____).
9. Qu'est-ce que chacun dans le parterre a fait avant de sortir? (_____).

Proofreading Test -- Part I.

DIRECTIONS: In the following passage a certain number of mistakes have been purposely made, one per line. Underline the word which you think wrong (form, position, tense, etc.), and write the correct form in the space to the right of the page.

Time allowed: 30 minutes

EXAMPLE:	1. Il a le vu avant d'entrer.	1 l'a vu
	2. Il attendit pour son ami.	2 ~~pour~~
	3. Il l'a vu avant entrant.	3 d'entrer

Proofreading Test--Form A

1.	Trois maisons brûlait au fond de la place.	1.	_____
2.	et une quatrième était menacé. Le feu avait	2.	_____
3.	pris dans le chantier d'une marchand de	3.	_____
4.	bois et de charbons que occupait le	4.	_____
5.	rez-de-chaussée d'un de les bâtiments. Il	5.	_____
6.	avait gagnée la boutique d'un marchand	6.	_____
7.	de couleurs, sise à côté. Le quantité	7.	_____
8.	des matérials inflammables, accumulés là	8.	_____
9.	faisait, de cette accident, un désastre.	9.	_____
10.	Une foule se pressa, grandissant d'heure	10.	_____
11.	en heure, dans lequel je m'enfonçai.	11.	_____
12.	On entendaient passer sur elle comme un	12.	_____
13.	frisson d'épouvante. Il y se mêlait	13.	_____
14.	l'appel des personnes quoi s'organisaient	14.	_____
15.	pour faisant la chaîne et le cri des comman-	15.	_____
16.	dements donnés par leur chefs aux pompiers	16.	_____
17.	et aux soldats. Les crépitements du incen-	17.	_____
18.	die dominaient tout. La flamme, avivé par	18.	_____
19.	un grande vent, était si forte qu'on y voy-	19.	_____
20.	ait comme en plein jour. Je me souvienne.	20.	_____
21.	En aidant de mon meilleur à passer des seaux	21.	_____
22.	d'eau, je regardais vers le cathédrale.	22.	_____
23.	Les moins détails des deux flèches aiguës	23.	_____
24.	se distinguaient pierre par pierre. Au	24.	_____
25.	cette clarté je pusse reconnaître, parmi	25.	_____

Appendix B

Illustrative Item Types and Comments

by

Paul B. Diederich

SEEING RELATIONSHIPS BETWEEN SENTENCES

DIRECTIONS. In each pair of sentences, mark the relationship of the second sentence to the first. Disregard whether the sentences are true or false.

Example: What is your name?
Christopher Columbus Langdell.

1 belief 2 fact 3 conclusion 4 proof

1 Thirty-six people who ate that dessert got sick.
Something in the dessert caused their illness.

1 belief 2 example 3 inference 4 fact

2 When I turned on the ignition, nothing happened.
The battery was probably dead.

1 evidence 2 observation 3 hypothesis 4 application

3 What is the weight of a cubic foot of water?
A cubic foot of water weighs 62.5 pounds.

1 fact 2 conclusion 3 definition 4 judgment

4 Why do you think he was the one who damaged our mailbox?
Because he has damaged several mailboxes in the past.

1 cause 2 proof 3 purpose 4 reason

5 He failed the examination.
He never finished an assignment during the course.

1 cause 2 effect 3 evidence 4 conclusion

6 All bodies fall with equal velocity in empty space.
When our astronauts dropped a feather and a hammer, they struck the moon at the same time.

1 natural law 2 proof 3 assumption 4 example

7 When the planets are in an unfavorable position, take no action.
The position of the planets influences human affairs.

1 fact 2 superstition 3 evidence 4 cause

8 No written language has originated south of the equator.
The great land-masses lie mainly north of the equator.

1 necessary condition 2 explanation 3 opinion 4 coincidence

9 In our society, women tend to live longer than men.
But women in business do not live longer than men in business.

1 contradiction 2 definition 3 effect 4 objection

10 Many requirements for the job were listed.
Mark thought that most of them were silly.

1 objection 2 opinion 3 contradiction 4 objection

Note. In logic, there are more exact technical names for some of these relationships. This exercise uses only names that are familiar to most high school students.. The number of the name nearest to the technically correct relationship has been underlined as the best of the four choices.

February 8, 1985 Paul B. Diederich

NOTE: Numbers before each blank refer to the same numbers in the key immediately below. For example, write in your answer to sentence 1. Suppose it is 1B. To find out if you are right, you have to look down at 3- in the key. There you find 3-1B. That keeps you from seeing the answer to the next sentence.

COMMON SENTENCE PATTERNS

1. Active

1A Subject verbs. This exercise teaches. Students learn.
1B Subject verbs object. Students answer the questions.
1C Subject verbs object object. The questions give them trouble.

Mark each sentence 1A, 1B, 1C (or 0 if none of these).

1. Boy meets girl. 3____
2. Does boy meet girl? 6____
3. Yes, he meets her. 8____
4. The boy whistles. 2____
5. She gives him a smile. 5____
6. He likes her. 9____
7. He calls her "Honey." 7____
8. Why does she go away? 1____
9. To get him to follow. 10____
10. He does not follow. 4____

Key 1-1A 2-1A 3-1B 4-1A 5-1C 6-1B 7-1C 8-1B 9-1B 10-0

2. Passive

2A Subject is verbed. Many questions are asked.
2B Subject is verbed by. The questions are answered by the students.
2C Subject is verbed object ± by. Students are then given the answers.

Mark each sentence 2A, 2B, 2C (or 0 if none of these).

1. He is not fooled by girls. 9___
2. He has been fooled before. 4___
3. He is given a smile. 5___
4. He just stands there. 10___
5. The girl is now caught. 1___
6. She has to come back. 6___
7. Greetings are exchanged. 8___
8. She is given compliments by him. 3___
9. She is flattered by compliments. 7___
10. Each is attracted by the other. 2___

Key: 1-2A 2-2B 3-2C 4-2A 5-2C 6-0 7-2B 8-2A 9-2B 10-0

3. Descriptive

3A Subject is noun. The answers are letters and numbers.
3B Subject is adjective. The questions are difficult.
3C Subject is adverb. The answers are below.

Mark each sentence 3A, 3B, 3C (or 0 if none of these).

1. He is a fine boy. 6___
2. She is a beautiful girl. 9___
3. They are in love. 7___
4. Love is wonderful. 1___
5. They are in a daze. 4___
6. They are crazy. 2___
7. They are out of their minds. 10___
8. Who cares? 3___
9. Love is not reasonable. 8___
10. Love is a very fine thing. 5___

Key: 1-3B 2-3B 3-0 4-3C 5-3A 6-3A 7-3C 8-3B 9-3A 10-3C

Basic terms used in both traditional and modern grammar:

1. active, passive, connective; verb, subject, object, complement; helping verb
2. phrase, clause (independent, subordinate, coordinate); simple, compound, complex
3. noun, pronoun, adjective, adverb, preposition, conjunction, article, interjection
4. singular, plural, possessive; number, case, person; agree, modify, apposition
5. tense, perfect; infinitive, participle, gerund; conditional, parenthetical

Other terms are used in ordinary senses (present, past, future; assertion, question, command; element, parallel) or can be defined by those listed above.

July, 1968
Princeton, New Jersey 08540

Paul B. Diederich
Educational Testing Service

GRAMMAR TEST. Your name_____ Grade___ Date_____

Teacher_____ School_____

DIRECTIONS: Encircle the number of the best answer to each question.

The test is based on one sentence:

I have a little shadow that goes in and out with me and what can be the use of him is more than I can see.

1. This sentence may be hard to read because one comma has been left out. Where would you put a comma to break up the sentence into two main parts?

1- After shadow.
2- After me.
3- After him.
4- After more.

2. What kind of sentence is this?

1- Simple.
2- Complex.
3- Compound.
4- Compound-complex

3. What is I have a little shadow?

1- The subject of the sentence.
2- The first independent clause.
3- The first subordinate clause.
4- The subject of him (line 3).

4. What is that goes in and out with me?

1- The first independent clause.
2- A subordinate clause, object of have.
3- A subordinate clause modifying shadow.
4- A subordinate clause modifying goes.

5. What is and?

1- A coordinating conjunction.
2- A subordinating conjunction.
3- A relative pronoun.
4- A preposition modifying what.

6. What is and what can be the use of him?

1- The second independent clause.
2- A subordinate clause modifying shadow.
3- A subordinate clause, subject of is.
4- A subordinate clause, object of see.

7. What is than I can see?

1- The second independent clause.
2- A subordinate clause, object of is.
3- A subordinate clause, object of more.
4- A subordinate clause modifying more.

8. What is is?

1- Verb of second independent clause.
2- Verb of second subordinate clause.
3- Verb modifying more.
4- A verb that does not have a subject.

9. What is more?

1- A coordinating conjunction.
2- A subordinating conjunction.
3- An adverb modifying than I can see.
4- A linking-verb complement.

10. What is the subject of the first independent clause?

1- I.
2- shadow.
3- I have a little shadow.
4- that goes in and out with me.

11. What is the subject of the second independent clause?

1- shadow.
2- that goes in and out with me.
3- what can be the use of him.
4- more than I can see.

12. How many subordinate clauses are there in this sentence?

1- One.
2- Two.
3- Three.
4- Four

13. What is the subject of the first subordinate clause?

1- shadow.
2- that.
3- what.
4- more.

NOTE. Tests like this have been shown to have low validity as indicators of writing ability, but of all the kinds of tests suggested in my book, Measuring Growth in English, this one was by far the most widely used. It tests only ability to name parts of a sentence in traditional terms, but this may have some value in its own right for highly verbal students.

Here is the sentence again:

I have a little shadow that goes in and out with me and what can be the use of him is more than I can see.

14. What is the subject of the second subordinate clause?

1- what.
2- use.
3- him.
4- more.

15. What is the subject of the third subordinate clause?

1- There is no third subordinate clause.
2- what.
3- use.
4- I.

16. What is the verb of the first independent clause?

1- have.
2- goes.
3- can be.
4- can see.

17. What is the verb of the second independent clause?

1- goes.
2- can be.
3- is.
4- can see.

18. What is shadow?

1- Subject of the whole sentence.
2- Object of have.
3- A linking-verb complement.
4- Object of the preposition little.

19. What are in and out?

1- Prepositions.
2- Adverbs.
3- Objects of goes.
4- Adjectives modifying with me.

20. What does with me modify?

1- shadow.
2- have.
3- goes.
4- in and out.

21. What is what?

1- A relative pronoun.
2- An interrogative pronoun.
3- An indefinite pronoun.
4- A personal pronoun.

22. What is of him?

1- Object of the verb use.
2- Prepositional phrase modifying use.
3- Prepositional phrase, subject of is more.
4- Prepositional phrase modifying can be.

23. What is than?

1- A coordinating conjunction.
2- A subordinating conjunction.
3- An adverb modifying can see.
4- A relative pronoun, object of can see.

24. Can be is a different form of the same verb as

1- have.
2- goes.
3- is.
4- can see.

25. What is can in can be and can see?

1- An adverb.
2- An auxiliary.
3- The subject.
4- The object.

26. The subordinate clauses in this sentence have three of the following functions. Which one do they not have?

1- Noun.
2- Verb.
3- Adjective.
4- Adverb.

Here is the sentence again: I have a little shadow that goes in and out with me and what can be the use of him is more than I can see.

Rewrite this sentence in as many of the following ways as you can. Use the same words that are in this sentence but change the form and order of these words as required. Try not to change or omit any of the ideas expressed by this sentence. Each rewitten version should be a single complete sentence.

27. Start with I had a little shadow.

28. Start with I cannot see the use.

29. Start with The children had.

30. Start with Do you have.

31. Start with What can be the use.

32. Start with Going in and out with me.

33. Start with More than I can see.

34. Start with Go in and out.

TESTS FOR READERS

In my study of the use of college-educated housewives as "lay readers" to help high school English teachers grade and correct papers (1958-59), we wanted to make sure that the selected housewives were highly intelligent, could tell a good paper from a poor one, could detect mistakes and correct them, and could write well themselves. For a rough indication of verbal aptitude we used a long, difficult vocabulary test. For ability to judge the general merit of student papers, we used four short papers on the same subject that skilled readers had placed at distinctly different levels of merit. For ability to spot and correct mistakes we used an actual student paper triple-spaced and asked them to insert corrections. For ability to write well themselves, we asked them to write a short paper on the advantages and disadvantages of using people other than teachers to help students improve their writing.

The tests would not have had to be highly valid in view of the high selection ratio. At least five times as many women applied for each position as we were able to use in every one of these 16 cities. In Bound Brook, the smallest, 64 women applied for four positions; in Detroit, 610 women had to be tested for the 16 positions then available. Those who were selected proved to be a splendid lot. By several independent criteria, many of them were more highly qualified than the teachers they assisted. Out of 54 readers in one school system, more of them had Master's degrees in English and more years of teaching experience than these teachers. This was due in part to the fact that many fine teachers had to take time out to have children but wanted part-time work at their former trade. It became standard practice to assign a veteran reader to a younger teacher and vice versa.

While the first tests were face-valid and yielded highly satisfactory candidates, there were two things wrong with them: they were too easy for the highly qualified women who applied, and they took much longer to score than was necessary for initial screening. It did not matter much if a few unsuitable candidates got by, because we always accepted at least twice as many as could actually be used, explaining that some would be used as alternates in case of sickness or family obligations of higher priority. After the first batch of papers was corrected and returned, it became obvious that some were not very good at this sort of work, or at least were thought to be no good by the old battle-axes they were supposed to assist. After a review of the evidence by the department head, sometimes the "old battle-ax" was denied the use of a reader; sometimes the reader was retired to the rank of "alternate" but was in fact not used except in emergencies.

To simplify and shorten the screening test and its scoring, we adopted the attached three tests of vocabulary, spelling, and mechanics (detection of errors in sentences). All three are made up of items found to be most difficult in large-scale studies, and evidence to that effect is given at the end of each test to reassure candidates who found them excessively difficult. The item-types used are not the most valid, but they go very fast and are valid enough for screening purposes. Spelling is not highly correlated with anything else in the verbal domain, but it is essential for readers. The "mechanics" test uses an item-type suggested by Don Melville that proved surprisingly valid even as an index of writing ability. The items chosen were the hardest out of 580 in this format used in a large-scale tryout. Many of them deal with "fine points" that even I regard as linguistic superstitions, but they are marked as errors by college teachers. Anything easier than this was answered correctly by nearly all candidates.

Name _____ School _____ Date _____

TESTS FOR READERS

A. VOCABULARY

DIRECTIONS: Put an X through R (right) if the underlined word is used correctly, or through W (wrong) if it is used incorrectly.

EXAMPLES

R ~~X~~ a. He accrued his money in real estate.

~~X~~ W b. She augmented her income by free-lance writing.

R W 1. He complained that the job was botched.

R W 2. The chamberlain was still cleaning my hotel room.

R W 3. The teacher consigned a topic for the next essay.

R W 4. They rode down the sharp declivity to the meadow below.

R W 5. He deprecated her efforts to find a new career.

R W 6. The duress of his appointment was one year.

R W 7. She did not buy the coat because the price was extraneous.

R W 8. His closing speech was a model of forensic oratory.

R W 9. The committee will investigate illicit traffic in drugs.

R W 10. His investitures in oil stocks left him bankrupt.

R W 11. When the mixture was overt, she put it in the oven.

R W 12. The tower was erected to perpetrate his memory.

R W 13. Plagiarism in student writing is usually obvious.

R W 14. She wrote detective stories under a pseudonym.

R W 15. The terms of the truce promised no reprisals.

R W 16. He had a long soliloquy with his lawyer.

R W 17. He accepted his sentence with stoical indifference.

R W 18. The star's tantamounts nearly drove his director crazy.

R W 19. She was too virtual to accept the suggestion.

R W 20. The aborigines of that island still practice cannibalism.

OVER

B. SPELLING

DIRECTIONS: Put an X through R (right) if the spelling is correct, or through W (wrong) if the spelling is incorrect.

EXAMPLES

X W a. acceptable
R X b. amature

R	W	1.	accomodate	R	W	26.	incidently
R	W	2.	achievment	R	W	27.	indispensible
R	W	3.	acquaint	R	W	28.	influencial
R	W	4.	actuality	R	W	29.	irrelevent
R	W	5.	adolesence	R	W	30.	liveliest
R	W	6.	alright	R	W	31.	magnificence
R	W	7.	arguement	R	W	32.	maintainance
R	W	8.	attendent	R	W	33.	maneuver
R	W	9.	authoratative	R	W	34.	Negroes
R	W	10.	catagory	R	W	35.	noticable
R	W	11.	characterised	R	W	36.	occurrence
R	W	12.	consistant	R	W	37.	playwrite
R	W	13.	contraversial	R	W	38.	primative
R	W	14.	dependant	R	W	39.	procede
R	W	15.	disastrous	R	W	40.	propogate
R	W	16.	devine	R	W	41.	psychosomatic
R	W	17.	dominent	R	W	42.	psycopathic
R	W	18.	embarrassment	R	W	43.	persue
R	W	19.	existance	R	W	44.	reminisce
R	W	20.	fantasies	R	W	45.	substancial
R	W	21.	falacy	R	W	46.	surpress
R	W	22.	fourty	R	W	47.	synonamous
R	W	23.	hinderance	R	W	48.	temperament
R	W	24.	hungrily	R	W	49.	tendancy
R	W	25.	hypocrit	R	W	50.	villain

Note: For college preparatory seniors, these words were the most difficult of 1,000 words tried out in this format.

C. MECHANICS

DIRECTIONS: Put an X through the number of the line that contains an error. If there is no error, put an X through 0.

EXAMPLES

A 1 Although I have never
X saw kangaroos, I know
3 what they look like.
0

B 1 She promised to return the
2 book within three weeks.
3 Has she kept her promise?
X

---

1 1 Smatherson was quite aggravated
2 by the numerous insults aimed
3 at him by his political opponent.
0

2 1 The illusion to Chaucer in
2 Duff's opening remarks was
3 proof of his wide reading.
0

3 1 "All right," she finally said.
2 "You may swim wherever you like
3 except at the old stone quarry."
0

4 1 When you have finished here,
2 bring your sketch to
3 Clancy, the art editor.
0

5 1 In comparison to our form
2 of government, the French
3 government is very complex.
0

6 1 Only twenty-one students
2 comprise the entire ninth
3 grade at our high school.
0

7 1 Wherever they looked,
2 a continual expanse of
3 desert met their gaze.
0

8 1 For what reasons did
2 your mother and father
3 immigrate from England?
0

9 1 Although Mike was nervous about
2 the test, he made less errors
3 than anyone else in the class.
0

10 1 What is your alibi for
2 not having done your
3 homework last night?
0

11 1 He inferred in his interview
2 that he would be quite willing
3 to accept the new position.
0

12 1 Of these three plans, I like the latter
2 most. It is the one that is most likely
3 to provide fun for everyone who attends.
0

13 1 We are liable to see some
2 faculty members at the
3 football game on Saturday.
0

14 1 I lay in bed all day
2 yesterday, but I am
3 still very tired.
0

15 1 No one is permitted to make
2 any noise when our teacher
3 is reading a poem out loud.
0

16 1 She told me to
2 plan on having
3 dinner with her.
0

17 1 Sally was criticized for
2 not being able to name the
3 principle rivers of Canada.
0

18 1 Steve has said that he will
2 come to the party, provided
3 that he may bring a friend.
0

ILLUSTRATIVE ITEMS FOR AN EXERCISE ON "SEEING RELATIONSHIPS"

DIRECTIONS: Mark each numbered sentence

C if it is a possible CAUSE of the event reported in the underlined statement

E if it is EVIDENCE (not necessarily conclusive) of the truth of the statement

R if it is a possible RESULT (or consequence) of the event but not acceptable as evidence of the truth of the statement (If it is both, mark it ER.)

N NONE OF THESE. Assume that all statements are true--at least as reports.

Jane stumbled over a dog.

1 Jane said that she stumbled over a dog.
2 The dog ran away on three legs.
3 Jane had dog hairs on her skirt later that day.
4 Jane almost never looks where she is going.
5 Jane is not always truthful.
6 The dog had been asleep on the sidewalk.
7 The dog belonged to a neighbor.
8 Jane's brother said that he saw her stumble over the dog.
9 Jane had a bruise on her knee that evening.
10 Jane should have been more careful.

The bulb in this lamp is burned out.

1 John knocked the lamp over.
2 When I turned the switch on, nothing happened.
3 The bulb is screwed in tightly.
4 The bulb has been in use two years.
5 Another bulb on this same circuit is lighted.
6 I have to buy a new light bulb.
7 Manufacturers are using poor materials in bulbs this year.
8 Everyone in this house leaves lights on when they are not needed.
9 When I shook the bulb, it tinkled.
10 This bulb has a burned-out filament.

Driver A was responsible for the accident.

1 There was no stop light at the intersection.
2 The whole side of car B was crumpled.
3 Inspection revealed that Driver A's brakes were faulty.
4 Driver A had to pay for the damage to Car B.
5 Driver B's tires left a black skid-mark on the pavement.
6 Driver A was nearsighted.
7 Driver B was driving on a through street.
8 Driver B had four passengers in his car.
9 Driver A did not see the stop sign at the intersection.
10 Driver B did not slow down before the intersection.

Other types of relationships:

hypothesis, conclusion, proof, argument, analogy, illustration, opinion, fact
observation, generalization, classification, instance, application, inference
natural law, moral law, statute law, norm, convention, fact
purpose, cause, necessary condition, sufficient condition, correlation
definition, distinction, assumption, postulate, axiom, theory
"is" of definition, identity, existence, location, description, assertion

February 8, 1985 Paul B. Diederich

A TEST OF UNDERSTANDING QUOTATIONS

Directions: On the answer sheet, put an X through the letter of the best of the three explanations of each quotation. For example:

0. We have met the enemy, and they are ours.
   A We have fought a battle, and we won.
   B We found the "enemy" but they turned out to be our own troops.
   C We are their enemies, and they are ours.

The answer sheet would be marked as follows: 0. ✕ B C because A is the best of the three explanations of this quotation.

You must not expect the best explanation to say exactly what you think the author meant. It need be only a bit closer to this meaning than the other two explanations. Do not make any marks on this copy of the test.

---

1. The story of any man's real experience finds its startling parallel in that of every one of us.
   A Truth is stranger than fiction.
   B Whenever we hear what really happened to someone else, it is surprising to find that something like it has happened to us.
   C Every person has interesting experiences that would make a good story if he only knew how to write about them.

2. Woman would be more charming if we could fall into her arms without falling into her hands.
   A Life would be more pleasant if we could go straight to our goal and not run into the obstacles that confront us every day.
   B A woman would be more desirable if she could be admired without being possessed.
   C Women would be more charming if we could love them without being dominated by them.

3. The thing generally raised on city land is taxes.
   A The only thing that grows on the land in cities is the amount of tax you have to pay.
   B Nothing is grown in cities because their taxes are too high.
   C It is too crowded in the city to raise any animals or crops.

4. He was not of an age but for all time.
   A His work is appreciated by young and old alike.
   B This person never seems to be old-fashioned.
   C His greatness will outlive the age in which he lived.

5. Those who live in glass houses should not throw stones.
   A You should keep your private affairs to yourself.
   B People who have faults of their own should not criticize faults in others.
   C Don't do things to other people that you would not want them to do to you.

6. There is only one thing in the world worse than being talked about, and that is not being talked about.
   A One may benefit from the criticism of other people.
   B We like to be talked about, if it is done in a friendly way.
   C It is better to be criticized than not even noticed.

NOTE. All answers were based on explanations of these sentences written by college freshmen. A large sample of such items was tried out in experimental forms of SAT and proved to be equal to conventional reading comprehension items in difficulty and discrimination but no better, so they were not adopted.

If he were alive today, the genial humorist Will Rogers would not have dared to say, "All I know is what I read in the papers." Yet, despite the message-massaging impact of new McLuhanic media, the newspaper continues to inform, entertain, and influence men. PAUL B. DIEDERICH, *Educational Testing Service, and* MARVIN MASKOVSKY, *American Newspaper Publishers Association Foundation, describe one of the most significant developments in the use of newspapers by students.*

CURRENT discussions of "the knowledge industry" focus upon the entry of industrial giants into the field of educational materials, marked by such acquisitions as Holt, Rinehart & Winston by CBS ($280 million), Ginn and Company by Xerox ($127 million), American Book Company by Litton ($71.4 million), and Science Research Associates by IBM ($62 million). Fear has been expressed that effective control of what is taught in our schools may soon rest in the hands of large corporations, and what can be sold at a profit may be a stronger determinant than what contributes to pupil growth or to the national welfare.

Rather than indulging in wild speculation, educators might do well to consider what has actually happened as a result of an earlier invasion of the educational establishment by a relatively unobtrusive segment of "the knowledge industry," the daily newspaper. Although newspapers had been used as instructional materials by interested teachers over a long period of time, there was no concerted program until 1956. Then the American Newspaper Publishers Association sponsored a meeting of newspaper publishers with representatives of the National Education Association, the National Council for the Social Studies, the National Council of Teachers of English, the National Association of Secondary School Principals, the American Association of School Administrators, and the Association for Supervision and Curriculum Development.

The purpose of this meeting was to consider how newspapers and educators could work together in pursuit of accepted educational objectives without regard to the special interests of newspapers. The outcome was a decision to study how newspapers were

then being used by a national sample of teachers, how widely they were used, and over what periods of time. The study, conducted by educators and supported by ANPA, reported in 1957 that existing uses of newspapers were inadequate in both methods and extent, but that large numbers of teachers were anxious to learn how to use newspapers more effectively.

Study then turned into action with the establishment of summer workshops for teachers, sponsored jointly by the American Newspaper Publishers Association and the National Council for the Social Studies. The first workshops were held in 1958 and continue today to offer graduate credit to selected teachers who are given scholarships by the local newspaper. Currently the ANPA Foundation and NCSS sponsor two-week workshops at Syracuse University, the University of Iowa, and the University of California at Los Angeles. Teachers who have attended these workshops now conduct at least 30 shorter state and regional workshops and seminars each year. Lectures, demonstrations, and discussions of ways of using newspapers for educational purposes are also conducted in hundreds of school districts throughout the country.

The "Newspaper in the Classroom" Program

The substantial and widespread use of newspapers that has resulted is known as the "Newspaper in the Classroom" program. At the last count it involved 322 newspapers, 17,600 schools, and 48,000 teachers. It reached over 3 million students with nearly 68 million copies of newspapers. Yet relatively few educators who are appalled at the invasion of their domain by "the knowledge industry" are aware that such a program exists!

The program has some central direction in that successful procedures developed by teachers are continually reported by educational representatives of newspapers to the ANPA Foundation and are passed along by the Foundation to other newspaper teachers throughout the country. Conferences of these representatives are held to share ideas about how to initiate, develop, and support better types of newspaper programs. Educational publications of the leading newspapers like "The American Newspaper" by Wallace Carroll and "Get More Out of Your Newspaper" by Theodore M. Bernstein, both published by *The New York Times*, are widely circulated. Hundreds of inquiries about various aspects of the program are answered by the Foundation staff.

Yet there is really no single "Newspaper in the Classroom" program. There are 322 different programs, each developed locally by interested teachers and administrators in cooperation with the local newspaper. Each program has some unique features. Although ideas gleaned from successful programs may be suggested by the local newspaper representative, and attention may be called to study guides and other instructional materials, all policy decisions are made by the local educators in charge of each program. It is the intention of the newspaper publishers to cooperate in these programs, not to dominate them. It costs them a good deal in time and money, but they would not remain in the newspaper business very long without a commitment to public service.

Such programs are directed toward a wide range of educational objectives but nearly all have the following objectives in common:

1. Competence in reading newspapers.
2. Understanding the role of newspapers in a free society.
3. Interest in what is going on and in what people say about it.

The Development of a Newspaper Test

After the program had developed for ten years, members of the ANPA Foundation came to Educational Testing Service to find ways to measure the attainment of these objectives. This long a delay is by no means unusual in an educational enterprise. Left to themselves, the teachers might have waited fifty years.

The most obvious means of measuring the effects of such a program was a current events test. This approach was rejected because there is no way of telling where a student's information and interest in news came from. It might be from radio or television newscasts, family conversation, class discussion, the weekly news publication written for students, the daily newspaper, or news magazines.

This difficulty imposed one inescapable requirement on any proposed measure. It would have to be tried out in a national sample of classes that had received newspaper training and in classes of equal reading ability in the same grades and schools that had never made any systematic use of adult newspapers. Only those questions would be selected for the published tests on which trained newspaper readers did better than the untrained. Questions that both groups answered equally well might be answered on the basis of general intelligence, general reading ability, general information, and access to other media. But if the two groups were equal in these respects,

questions answered by a larger percentage of trained newspaper readers would show the effects of their training.

The first objective to be measured, competence in reading newspapers, suggested a test based on simulated newspapers, similar to daily newspapers in all respects except length and coverage of events that could be dated. Since students would have to find and read the material to which the questions referred and then answer the questions in a class period of about 40 minutes, a short newspaper would be all that they could handle. The events treated would all be imaginary but should be of a kind that might occur at any time, since the published tests would first be used some 18 months after the material was written and would remain in use from five to ten years before a major revision was necessary.

A further requirement was that the simulated newspapers should be written by skilled journalists in the style of the best daily newspapers. They should not be "written down" to the supposed capacity of high school students because one of the main points of the test was to find out how well these students could deal with material written for adults.

The second objective, understanding the role of newspapers in a free society, could be tested by questions dealing with the rights and responsibilities of newspapers, their reasons for publishing the various types of articles, the credibility of their reports, the possibilities of distortion or suppression, and the like. It was expected that relatively few questions of these kinds would survive the tryout, since this objective is approached in so many different ways that a national sample of newspaper classes would probably do better than the control classes in only a few types of questions. These would give the objective at least token recognition that might be strengthened in subsequent editions as the tests became widely adopted. They would also show teachers how to test the particular understandings they were trying to develop in their newspaper classes.

The third objective, interest in what is going on and in what people say about it, could hardly be tested at all in a test based on simulated newspapers. Students who had this interest might show it by their familiarity with the various types of newspaper material and their success in answering questions that require background and experience in reading newspapers. More direct evidence might be supplied by teacher-made tests on events treated in the local newspaper during the period of instruction.

Once these guidelines were adopted, the ANPA Foundation agreed to support the development of a newspaper test for junior and senior high schools in cooperation with Educational Testing Service. The Test Development Committee was headed by Dr. John H. Haefner, Professor of Social Studies Education at the University of Iowa and past president of the National Council for the Social Studies, which lent its sponsorship to the undertaking. The test editor was Dr. Paul B. Diederich of Educational Testing Service, assisted by Dr. Harry D. Berg, Office of Evaluation Services, Michigan State University, and editor of the NCSS Yearbook, *Evaluation in Social Studies* (1965). The test questions were reviewed by a Teachers Advisory Panel headed by Mrs. Virginia Riggs of Ector High School, Odessa, Texas.

The Foundation itself assumed responsibility for the production of two simulated newspapers (one for pretests, the other for posttests) of four pages each, tabloid size. Each has at least three items of local news and one of state, national, and international news; editorials on two of these imaginary events; a column of commentary on a perennial political issue; an editorial cartoon (but no comics); an opinion poll; a review of a play or art exhibit by local talent; a story of an event in which teenagers were involved; a sports story; a personal advice column; letters to the editor; news photographs and a few advertisements. Material that might be offensive to teachers, parents, and minority groups was avoided.

Four teachers with experience in item writing were then employed during the summer of 1968 to write questions on these newspapers testing the kinds of skill, knowledge, and judgment that intelligent newspaper reading requires. They produced nearly 700 questions, most of them with a choice of four answers, and 400 of these were tried out in January 1969 in 61 junior and senior high schools in all parts of the country. These schools were selected by the Foundation with the help of educational representatives of newspapers. For each newspaper class that was tested, the teacher and principal selected a control class in the same grade with equal scores on reading comprehension tests but without newspaper training. It is well known to researchers that if reading comprehension scores are equal in a large national sample of classes, then general intelligence, family background, and access to other media are also likely to be equal. There were eight tryout forms of 50 items each, and these were arranged before shipment in a "spiral" order so that the first student in each class would get Form 1, the next Form 2, and so on. Thus every class tested received all eight tryout forms but only an eighth of the students took each form.

Results of Newspaper Testing

Items selected for the senior high tests were answered correctly by an average of 66 percent of the senior high newspaper students and 57 percent of the control students. For junior high tests the corresponding figures were 51 percent of the junior high newspaper students and 44 percent of the control students. Since the items that best showed the effects of newspaper training in junior high schools were more difficult than is desirable in a test, they have since been rewritten in simpler words without changing their point.

Every item used in the published tests showed an advantage of newspaper students over control students. Hence it can be asserted that the tests measure something attributable to newspaper training—not to general intelligence, reading ability, home background, or access to other media. This claim is supported by a classification of the published items in 28 types of competence, such as ability to judge why certain statements were made, to distinguish what the newspaper said from what certain people were quoted as saying, to anticipate what would happen next in a given situation, and to recognize the kinds of language that would be used in different types of material. In all of these it is reasonable to suppose that newspaper training would make a difference.

Some of the publishers who supported the development of these tests hoped that the very tryout of questions would provide conclusive evidence of the value of the "Newspaper in the Classroom" program. They were told that the tryout was comparable only to the calibration of a yardstick that would later be applied to various types of programs to measure their relative effectiveness. Even then the yardstick could hardly show that all such programs were good—only that some were better than others. Studies of programs that consistently produced large gains from pretest to posttest would reveal procedures that other programs would be glad to adopt.

Although the hope of finding conclusive evidence of the value of all sorts of newspaper programs at this time was quite properly dismissed as premature, there is this much to be said for it. When the 400 questions were tried out in 61 schools, there was a real possibility that no differences between newspaper and control classes would be found.

It was gratifying, therefore, to find plenty of items showing differences in favor of newspaper students to make up a pretest and posttest at exactly the desired level of difficulty for senior high schools and at something close to it for junior high schools. It is unusual to be able to provide four tests in a hitherto untested field in which every item has shown a difference in the desired direction between the criterion groups. Although any one of these differences might have arisen by chance, the combined differences on the 30 items selected for each published form cannot reasonably be attributed to chance. In groups of the size used in school experiments, the probability is less than one in a hundred that differences as large as those found in the tryout might occur by chance.

For this reason, if Superintendents of Schools in cities that do not have a newspaper program order these tests and compare their average scores with those found in cities that have long-established programs, most of them will probably conclude that such a program is needed.

These differences are likely to increase as these tests are widely adopted. At present, newspaper teachers may be pursuing as many as 100 different goals with varying success. They may use foreign news to teach geography, editorials to teach expository writing, or stock market reports to teach arithmetic. Hence differences in the kinds of skill, knowledge, and judgment revealed by the tests may be largely accidental by-products of many different types of newspaper instruction. The tests will provide a common goal for all newspaper programs that will be widely accepted, and they define this goal more completely than most goals that teachers now have in mind. With this kind of target to shoot at, the superiority already shown by newspaper classes my be expected to increase. To that end, the test committee has suggested an interesting method of conducting class discussions of pretest items on which there was a substantial difference of opinion. Teachers who adopt this method and like it should be able to devise similar questions on the newspaper used in class.

Although teachers are usually reluctant to direct their teaching toward a published test, the kinds of skill, knowledge, and judgment demanded and revealed by this test are of crucial importance. They define more clearly than ever before the aims of most newspaper programs as revealed by their results, and they are close to the heart of education for citizenship. The effects of such tests on teaching can hardly fail to be good.

For information on the ANPA Foundation Newspaper Test and specimen sets, write: *Cooperative Tests and Services, Educational Testing Service, Princeton, New Jersey 08540.*

Reprinted from SOCIAL EDUCATION, Volume XXXIV, Number 2
February, 1970

Evaluating Innovations by Innovations in Evaluation

—PAUL B. DIEDERICH
Educational Testing Service, Princeton, New Jersey

In visiting interns with evaluation problems, I have been delighted to find that our high schools are moving again, but their testing, grading, record-keeping, and reporting practices are exactly what they were in 1935, when I began visiting the thirty schools of the Eight Year Study. Why are schools so conservative in measurement when they are innovative in everything else? My own conclusion is that trying to get individual teachers to do anything intelligent about measurement—and to keep on doing it—is a lost cause.

I see no hope for any significant improvement until responsibility for measurement of the four, five, or six most important objectives in each field is transferred from individual teachers to the department or team. Each department may also be given special responsibility for one objective of school-wide concern: social studies for democratic behavior, English for independent reading and study, science and math for critical thinking, guidance for academic interests, and so on. These objectives are now unmeasured or inadequately measured because what is everyone's responsibility becomes no one's responsibility.

Such a program is directed by an Evaluation Committee composed of heads of departments and special services, such as the library, guidance, student activities, and sports. This committee may meet only four times a year but each time for a full morning, with substitutes hired to cover classes. A clean break with the individualistic tradition of school measurement cannot be made by tired people who always have to meet after school. To avoid having to grapple with too many diverse objectives at the same time, the committee may start with just three fields of study and add three more fields per year.

In the public setting of this committee, each department head answers three basic questions: What major objectives will your department try to measure this year? By what sort of instrument or procedure? Within what dates? It is understood that no department will try to measure more than six major objectives (although these may have subheads) and may start with one or two. It is also understood that each department will report briefly at the next meeting of the committee on each measure that it administers. These reports usually include an analysis of the objective, a description of the measure, and a summary of results with conclusions and recommendations.

When this amount of pressure is applied, each department starts to measure one or more of its objectives, and these measures are usually superior to those devised by individual teachers for several reasons:

- First, they have to focus on major objectives because these are all that classes taught by different teachers have in common.
- Second, the questions, items, or tasks prepared by each teacher are reviewed by others.
- Third, it seems natural in the departmental setting to have written answers or other products rated independently by two teachers, and when the ratings differ by more than a certain amount, to have a third rating by a committee of the most experienced judges.
- Fourth, since a report on each measure is expected, it is necessary to make some kind of analysis or classification of the responses in order to discover the strengths and weaknesses of the instructional program.
- Fifth, the conclusions and recommendations of each report usually lead to changes in the program, and further measurement is expected to show the results of these changes.
- Sixth, since this whole process is reported to a committee of department heads, more thought and care are given to it than to measures that no one sees except the teacher, and no one asks about results, conclusions, or subsequent action.

An account of an experimental tryout of this program is given in Chapters 6, 7, and 8 of the 1967 Yearbook of the Association for Supervision and Curriculum Development (ASCD), *Evaluation as Feedback and Guide*, edited by Fred T. Wilhelms and obtainable at the ASCD Office, 1201 Sixteenth Street, N.W., Washington, D.C. 20036.

From DESIGN FOR LEADERSHIP: NASSP Internship Project News Report, April 1967. Washington, D. C.: National Association of Secondary-School Principals.